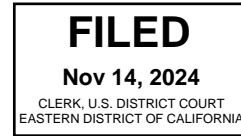


CAHILL GORDON & REINDEL LLP
JOEL KURTZBERG (*pro hac vice pending*, SBN NY 1758184)
FLOYD ABRAMS (*pro hac vice pending*, SBN NY 2835007)
JASON ROZBRUCH (*pro hac vice pending*, SBN NY 5753637)
32 Old Slip
New York, New York 10005
Phone: 212-701-3120
Facsimile: 212-269-5420
jkurtzberg@cahill.com



DOWNEY BRAND LLP
WILLIAM R. WARNE (SBN 141280)
MEGHAN M. BAKER (SBN 243765)
621 Capitol Mall, 18th Floor
Sacramento, CA 95814
Phone: 916-444-1000
Facsimile: 916-520-5910

Attorneys for Plaintiff X Corp.

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF CALIFORNIA
SACRAMENTO DIVISION

X CORP.,

Plaintiff,

v.

ROBERT A. BONTA, Attorney
General of California, in his
official capacity, and
SHIRLEY N. WEBER, Secretary of
State of California, in her
official capacity,

Defendants.

2:24-cv-2527 JAM CKD

Case No. ~~2:24 cv 3162 JAM CKD~~

**COMPLAINT FOR DECLARATORY AND
INJUNCTIVE RELIEF**

Plaintiff X Corp., by and through its attorneys, Cahill Gordon
& Reindel LLP and Downey Brand LLP, alleges for its complaint
against the above-named Defendants, as follows:

COMPLAINT

NATURE OF THE ACTION

1
2 1. Plaintiff X Corp. brings this action challenging the
3 constitutionality and legal validity of California Assembly Bill
4 No. 2655 ("AB 2655"), which is codified in law at Cal. Elec. Code
5 §§ 20510–20520.

6 2. AB 2655 requires large online platforms like X, the
7 platform owned by X Corp. (collectively, the "covered platforms"),
8 to remove and alter (with a label) – and to create a reporting
9 mechanism to facilitate the removal and alteration of – certain
10 content about candidates for elective office, elections officials,
11 and elected officials, of which the State of California disapproves
12 and deems to be "materially deceptive." It has the effect of
13 impermissibly replacing the judgments of covered platforms about
14 what content belongs on their platforms with the judgments of the
15 State. And it imposes liability on the covered platforms to the
16 extent that their judgments about content moderation are
17 inconsistent with those imposed by the State. AB 2655 thus violates
18 the First and Fourteenth Amendments of the United States
19 Constitution; the free speech protections of Article I, Section 2,
20 of the California Constitution; and the immunity provided to
21 "interactive computer services" under Section 230 of the
22 Communications Decency Act, 47 U.S.C. § 230(c).

23 3. Worse yet, AB 2655 creates an enforcement system that
24 incentivizes covered platforms to err on the side of removing

1 and/or labeling any content that presents even a close call as to
2 whether it is "materially deceptive" and otherwise meets the
3 statute's requirements. This system will inevitably result in the
4 censorship of wide swaths of valuable political speech and
5 commentary and will limit the type of "uninhibited, robust, and
6 wide-open" "debate on public issues" that core First Amendment
7 protections are designed to ensure. *New York Times v. Sullivan*,
8 376 U.S. 254, 270 (1964). As the United States Supreme Court has
9 recognized, our strong First Amendment protections for such speech
10 are based on our nation's "profound national commitment" to
11 protecting such debate, even if it often "include[s] vehement,
12 caustic, and sometimes unpleasantly sharp attacks on government
13 and public officials." *Id.*

14 4. AB 2655's problematic enforcement system provides
15 expedited causes of action for injunctive and other equitable
16 relief to the California Attorney General, every California
17 district attorney, every California city attorney, and to
18 candidates for elective office, elections officials, and elected
19 officials, to force covered platforms to remove certain "materially
20 deceptive content," alter that content, and comply with the
21 statute's reporting requirement. Even if the covered platform has
22 a robust process for investigating reported content, it will be
23 subject to such lawsuits for injunctive relief if it does not
24 remove or label the reported content within 72 hours. Enforcement

1 actions may be brought for "injunctive or other equitable relief
2 against any large online platform" to remove or label content that
3 should have been removed or labeled under the statute. See
4 §§ 20515(b), 20516. In short, covered platforms may be sued if
5 governmental officials or candidates think they have not censored
6 or labeled enough content; but the platforms may not be sued by
7 anyone if they have arguably censored or labeled too much content
8 under the statute. The result is a system that highly incentivizes
9 covered platforms to remove or label any content that presents a
10 close call to avoid lawsuits altogether.

11 5. AB 2655 suffers from a compendium of serious First
12 Amendment infirmities. Primary among them is that AB 2655 imposes
13 a system of prior restraint on speech, which is the "most serious
14 and the least tolerable infringement on First Amendment rights."
15 *Nebraska Press Ass'n v. Stuart*, 427 U.S. 539, 559 (1976). The
16 statute mandates the creation of a system designed to allow for
17 expedited "take downs" of speech that the State has targeted for
18 removal from covered platforms in advance of publication. The
19 government is involved in every step of that system: it dictates
20 the rules for reporting, defining, and identifying the speech
21 targeted for removal; it authorizes state officials (including
22 Defendants here) to bring actions seeking removal; and, through
23 the courts, it makes the ultimate determination of what speech is
24 permissible. Rather than allow covered platforms to make their

1 own decisions about moderation of the content at issue here, it
2 authorizes the government to substitute its judgment for those of
3 the platforms.

4 6. It is difficult to imagine a statute more in conflict
5 with core First Amendment principles. As the United States Supreme
6 Court has held, "it is a central tenet of the First Amendment that
7 the government must remain neutral in the marketplace of ideas."
8 *Hustler Magazine, Inc. v. Falwell*, 485 U.S. 46, 56 (1988). Even
9 worse, AB 2655's system of prior restraint censors speech about
10 "public issues and debate on the qualifications of candidates," to
11 which the "First Amendment affords the **broadest protection**" to
12 ensure the "unfettered interchange of ideas for the bringing about
13 of political and social changes desired by the people." *McIntyre*
14 *v. Ohio Elections Comm'n*, 514 U.S. 334, 346 (1995).¹

15 7. AB 2655 imposes a prior restraint on speech because it
16 provides, pursuant to Sections 20515(b) and 20516, expedited causes
17 of action under Section 35 of the California Code of Civil Procedure
18 through which political speech can be enjoined before there occurs
19 a "final judicial determination" that the "speech is unprotected."
20 *Isaksen v. Mazu Publ'g Co.*, 2005 WL 8176605, at *3 (E.D. Cal. Mar.
21 29, 2005) (citing *Vance v. Universal Amusement Co.*, 445 U.S. 308
22 (1980)) (denying motion for preliminary injunction as to already
23

24 ¹ Unless otherwise indicated, emphases in quotes are added and internal citations
and quotations are omitted.

published speech because it would have constituted a prior restraint). Although the statute tasks plaintiffs with demonstrating “through clear and convincing evidence” – see §§ 20515(b), 20516) – that the speech is “materially deceptive” content that otherwise meets the statute’s requirements, that showing **does not** amount to proof that the speech is constitutionally unprotected. See *Kohls v. Bonta*, 2024 WL 4374134, at *3-5 (E.D. Cal. Oct. 2, 2024) (holding that a companion statute, AB 2839, that provides a cause of action against individuals who post “materially deceptive content” – defined nearly identically as it is in AB 2655 – likely violated the First Amendment on its face because the statute’s “legitimate sweep pales in comparison to the substantial number of its applications . . . which are plainly unconstitutional”); see also *Garcia v. Google, Inc.*, 786 F.3d 733, 747 (9th Cir. 2015) (forcing Google through “takedown order” to remove content previously published on YouTube prior to a final determination that the content was unprotected amounted to a “classic prior restraint on speech”); *Living Vehicle, Inc. v. Kelley*, 2023 WL 2347442, at *9 (C.D. Cal. Jan. 20, 2023) (citing *Alexander v. United States*, 509 U.S. 544, 550 (1993); *Garcia*, 786 F.3d at 746-47) (prior restraints “refer either to injunctions that restrict future speech or require takedowns of currently-published speech”); *SolarPark Korea Co. v. Solaria Corp.*,

1 2023 WL 4983159, at *11 (N.D. Cal. Aug. 2, 2023) (same), *appeal*
2 *dismissed*, 2023 WL 9860831 (9th Cir. Sept. 28, 2023).

3 8. Further evidencing that AB 2655 imposes a prior restraint
4 on speech is that, apart from the expedited suits for injunctive
5 and other relief authorized under Sections 20515(b) and 20516, (i)
6 nothing in AB 2655 prevents the enjoinder of speech through a
7 temporary restraining order or preliminary injunction alternative
8 to or in addition to such suits; (ii) AB 2655 mandates the immediate
9 removal of speech, without a determination that it is unprotected,
10 so long as it is "substantially similar" to speech "previously
11 removed" under the statute, § 20513(c); and (iii) the statute acts
12 as an overarching prior restraint by, in its pursuit of eliminating
13 certain speech altogether, imposing a system of censorship that
14 requires covered platforms that wish to avoid being sued to block
15 speech within 72 hours absent a final ruling that the speech is
16 unprotected.

17 9. Even if AB 2655 were not a prior restraint, it still
18 violates the First Amendment because it runs counter to the United
19 States Supreme Court's recent decision in *Moody v. NetChoice, LLC*,
20 in which the Court held, in no uncertain terms, that when a social
21 media platform "present[s] a curated and 'edited compilation of
22 [third party] speech,'" that presentation "is itself protected
23 speech." 144 S. Ct. 2383, 2409 (2024) (quoting *Hurley v. Irish-*
24 *Am. Gay, Lesbian & Bisexual Grp. of Boston*, 515 U.S. 557, 570

(1995)); see also *id.* at 2401 ("A private party's collection of third-party content into a single speech product (the operators' 'repertoire' of programming) is itself expressive, and intrusion into that activity must be specially justified under the First Amendment."); *id.* at 2405 (quoting *Miami Herald Pub. Co. v. Tornillo*, 418 U.S. 241, 258 (1974)) ("'The choice of material,' the 'decisions made [as to] content,' the 'treatment of public issues' – 'whether fair or unfair' – all these 'constitute the exercise of editorial control and judgment.' . . . **For a paper, and for a platform too.**"). Because AB 2655 impermissibly replaces the judgments of the covered platforms about what speech may be permitted on their platforms with those of the government, it cannot be reconciled with the Supreme Court's decision in *Moody*.

10. AB 2655 disregards numerous significant First Amendment holdings by the Supreme Court in *Moody* – specifically, that (i) it is not a "valid, let alone substantial" interest for a state to seek "to correct the mix of speech" that "social-media platforms present," *id.* at 2407; (ii) a "State 'cannot advance some points of view by burdening the expression of others,'" *id.* at 2409 (quoting *Pac. Gas & Elec. Co. v. Pub. Utilities Comm'n of California*, 475 U.S. 1, 20 (1986)); (iii) the "government may not, in supposed pursuit of better expressive balance, alter a private speaker's own editorial choices about the mix of speech it wants to convey," *id.* at 2403; (iv) "it is no job for government to

1 decide what counts as the right balance of private expression – to
2 ‘un-bias’ what it thinks biased, rather than to leave such
3 judgments to speakers and their audiences. That principle works
4 for social-media platforms as it does for others,” *id.* at 2394;
5 and (v) “[h]owever imperfect the private marketplace of ideas,” a
6 “worse proposal” is “the government itself deciding when speech
7 [is] imbalanced, and then coercing speakers to provide more of some
8 views or less of others,” *id.* at 2403.

9 11. AB 2655 also runs counter to the First Amendment’s
10 staunch protection of core political speech. By imposing
11 unintelligible prohibitions on allowing a specific category of
12 speech under threat of enormous liability if it is not labeled
13 and/or removed to the government’s satisfaction, AB 2655 “acts as
14 a hammer instead of a scalpel,” *Kohls*, 2024 WL 4374134, at *8,
15 greatly incentivizing covered platforms to censor *all* content that
16 could reasonably fall within the statute’s purview to avoid
17 substantial enforcement costs. This, in turn, will severely chill
18 important political speech – specifically, the use of exaggerated
19 or unfavorable visual means to undermine and combat political
20 opponents, which, as the Supreme Court has recognized, is ingrained
21 in the historical fabric of U.S. political commentary and subject
22 to the strongest of First Amendment protections.

23 12. Whether it be “Walt McDougall’s characterization” in 1884
24 “of Presidential candidate James G. Blaine’s banquet with the

1 millionaires at Delmonico's as 'The Royal Feast of Belshazzar'" or
2 contemporary imaginings of Donald Trump's arrest² or what a second
3 term under President Biden would look like,³ "graphic depictions
4 and satirical cartoons have played a prominent role in public and
5 political debate," and "it is clear that our political discourse
6 would [be] considerably poorer without them." *Falwell*, 485 U.S.
7 at 54-55. Indeed, "YouTube videos, Facebook posts, and X tweets
8 are the newspaper advertisements and political cartoons of today,
9 and the First Amendment protects an individual's right to speak
10 regardless of the new medium these critiques may take." *Kohls*,
11 2024 WL 4374134, at *5. Contemporary commentators frequently use
12 artificial intelligence to generate this type of valuable
13 commentary. *Id.*

14 13. There is a long history of the strongest of First
15 Amendment protections for speech critical of government officials
16 and candidates for public office that includes tolerance for
17 potentially false speech made in the context of such criticisms.
18 And there is a long history of skepticism of any governmental
19 attempts to regulate such content, no matter how well-intentioned
20 they may be. As both the Supreme Court and Judge Learned Hand have

21 ² Ex. 1 (Eliot Higgins (@EliotHiggins), X (Mar. 20, 2023, 5:22 PM), formerly
22 available at <https://x.com/EliotHiggins/status/1637927681734987777> (last
visited Nov. 5, 2024)).

23 ³ Ex. 2 (GOP, *Beat Biden*, YouTube (Apr. 25, 2023),
24 <https://www.youtube.com/watch?v=kLMMxgtxQ1Y> (last visited Nov. 14, 2024)); see
also Ex. 3 (S. Comm. on Judiciary, Analysis of Bill No. AB 2655, 2023-2024 Reg.
Sess. (Cal. June 28, 2024)) at 7, 9 (citing this video as an example of how
"generative AI can spread misinformation regarding elections with ease").

1 noted, "[t]he First Amendment" "presupposes that right conclusions
2 are more likely to be gathered out of a multitude of tongues than
3 through any kind of authoritative selection. To many, this is,
4 and always will be, folly; but we have staked upon it our all."
5 *Sullivan*, 376 U.S. at 270 (quoting *United States v. Associated*
6 *Press*, 52 F. Supp. 362, 372 (S.D.N.Y. 1943) (Hand, J.)). AB 2655
7 runs counter to these principles by attempting to impose by
8 "authoritative selection" the permissible content on covered
9 platforms, rather than allowing the "multitude of tongues" engaging
10 in political debate and commentary on those platforms to do so.
11 See also, e.g., *Beilenson v. Superior Ct.*, 44 Cal. App. 4th 944,
12 954 (1996) ("Hyperbole, distortion, invective, and tirades are as
13 much a part of American politics as kissing babies and distributing
14 bumper stickers and pot holders. Political mischief has been part
15 of the American political scene since, at least, 1800. . . . 'Once
16 an individual decides to enter the political wars, he subjects
17 himself to this kind of treatment. . . . [D]eeply ingrained in our
18 political history is a tradition of free-wheeling, irresponsible,
19 bare knuckled, Pier 6, political brawls.'").

20 14. Accordingly, AB 2655 violates the First Amendment of the
21 United States Constitution and Article I, Section 2, of the
22 California Constitution, both facially and as-applied to X Corp.
23 AB 2655 imposes a prior restraint on speech that forces platforms
24 to censor only certain election-related content of which the State

1 of California disapproves and also directly and impermissibly
2 interferes with the constitutionally protected content-moderation
3 speech rights of covered social media platforms, like X. And AB
4 2655 does so notwithstanding that less speech-restrictive
5 alternatives would serve California's interest in protecting its
6 free and fair elections.

7 15. AB 2655 also directly contravenes the immunity provided
8 to the covered platforms by 47 U.S.C. §§ 230(c)(1) and 230(c)(2),
9 which prohibit (i) treating interactive computer service providers
10 as the "publisher or speaker of any information provided by another
11 information content provider," § 230(c)(1); and (ii) liability "on
12 account" of "any action" "taken to enable or make available
13 to information content providers or others the technical means to
14 restrict access to [objectionable] material," § 230(c)(2)(B).

15 16. First, in violation of § 230(c)(1), by providing causes
16 of action for "injunctive or other equitable relief against" the
17 covered platform to remove or (by adding a label) to alter certain
18 content posted on the platform by its users (see §§ 20515(b),
19 20516), AB 2655 treats covered platforms "as the publisher or
20 speaker of information provided by another information content
21 provider." 47 U.S.C. § 230(c)(1).

22 17. Second, in violation of § 230(c)(2)(B)'s prohibition on
23 holding platforms liable for "action[s] taken to enable or make
24 available to information content providers or others the technical

1 means to restrict access to [objectionable] material," AB 2655
2 provides causes of action for "injunctive or other equitable relief
3 against" covered platforms that attempt to comply with the
4 statute's reporting requirement, but do so in a manner that, in
5 the government attorney's view, does not meet the reporting
6 "require[ments]" of "subdivision (a) of Section 20515." § 20516.
7 In other words, a covered platform's attempt to comply with the
8 statute's reporting requirement (i.e., by creating a reporting
9 requirement for users to report content covered by the statute) is
10 an action, as contemplated by § 230(c)(2)(B), to make available
11 the technical means to restrict access to objectionable content,
12 and, in contravention thereof, AB 2655 imposes liability on any
13 covered platform that takes such action in a manner deemed
14 insufficient by the California government.

15 18. So too does AB 2655 violate the First and Fourteenth
16 Amendments of the United States Constitution for vagueness. AB
17 2655's requirements are so vague and unintelligible that covered
18 platforms cannot understand how to comply with them; thus, those
19 subject to its language will be compelled to over-censor speech to
20 avoid costly litigation over countless judgment calls surrounding
21 whether the statute prohibits particular pieces of content.

22 19. In pursuing this action, X Corp. seeks declaratory relief
23 and preliminary and permanent injunctive relief on the grounds that
24 AB 2655 (i) violates the free speech rights of X Corp. and the

1 other covered platforms under the First Amendment of the United
2 States Constitution and Article I, Section 2, of the California
3 Constitution, both facially and as-applied to X Corp.; (ii)
4 directly conflicts with, and is thus preempted by, the immunity
5 afforded to X Corp. by 47 U.S.C. §§ 230(c)(1) and 230(c)(2); and
6 (iii) violates the First and Fourteenth Amendments of the United
7 States Constitution because its requirements are so vague and
8 unintelligible that the covered platforms cannot understand what
9 they permit and what they prohibit, which will lead to blanket
10 censorship, including of valuable political speech.

11 20. In pursuing this action, X Corp. seeks to vindicate the
12 deprivation of constitutional rights under color of state statute,
13 ordinance, regulation, custom, and/or usage. X Corp. is also
14 entitled to attorneys' fees and costs if it prevails on any of its
15 § 1983 claims. See 42 U.S.C. § 1988.

16 **PARTIES**

17 21. Plaintiff X Corp. is a corporation organized and existing
18 under the laws of the State of Nevada, with its principal place of
19 business in Bastrop, Texas. X Corp. provides the X service, which
20 is a real-time, open, public conversation platform, where people
21 can see every side of a topic, discover news, share their
22 perspectives, and engage in discussion and debate. X allows people
23 to create, distribute, and discover content and has democratized
24 content creation and distribution. X allows users to create and

1 share ideas and information instantly through various product
2 features, including public posts.

3 22. AB 2655 applies to X Corp. because X is a "large online
4 platform," as defined by the statute – i.e., a "public-facing
5 internet website," "video sharing platform," and "social media
6 platform as defined by Section 22675 of the Business and
7 Professions Code"⁴ that "had at least 1,000,000 California users
8 during the preceding 12 months." § 20512(h).

9 23. Defendant Robert Bonta is the Attorney General of the
10 State of California and is charged with enforcing AB 2655. X Corp.
11 sues Attorney General Bonta in his official capacity as the person
12 charged with enforcing AB 2655.

13 24. Defendant Shirley Weber is the Secretary of State of the
14 State of California and is also charged with enforcing AB 2655. X
15 Corp. sues Secretary Weber in her official capacity as the person
16 charged with enforcing AB 2655.

17 JURISDICTION

18 25. This Court has jurisdiction over X Corp.'s federal
19 claims pursuant to 28 U.S.C. §§ 1331 and 1343(a) and 42 U.S.C.

20 ⁴ X is a "social media platform," as defined by Section 22675 of the Business
21 and Professions Code, because it is a public internet-based service or
22 application with users in California and (i) "[a] substantial function of the
23 service or application is to connect users in order to interact socially with
24 each other within the service or application" and (ii) it allows its users to
(a) "construct a public or semipublic profile for purposes of signing into and
using the service or application"; (b) "[p]opulate a list of other users with
whom an individual shares a social connection within the system"; and (c)
"[c]reate or post content viewable by other users, including but not limited to,
on message boards, in chat rooms, or through a landing page or main feed that
presents the user with content generated by other users."

1 § 1983, because X Corp. alleges violations of its rights under the
2 Constitution and laws of the United States. The Court has
3 jurisdiction over X Corp.'s state claim pursuant to 28 U.S.C.
4 § 1367.

5 26. This Court has authority to grant declaratory and
6 injunctive relief under the Declaratory Judgment Act, 28 U.S.C.
7 §§ 2201, 2202, and under the Court's inherent equitable
8 jurisdiction.

9 VENUE

10 27. Venue is proper in this Court under 28 U.S.C.
11 §§ 1391(b)(1) and 1391(b)(2) because the Defendants are located,
12 reside, and have offices in this judicial district and in the State
13 of California, and the violations of X Corp.'s rights are occurring
14 and will occur within this judicial district. AB 2655 was also
15 enacted in this judicial district.

16 FACTUAL ALLEGATIONS

17 **I. AB 2655's Statutory Scheme**

18 28. AB 2655, which applies to "large online platform[s],"
19 including "public-facing internet website[s]," "video sharing
20 platform[s]," and "social media platform[s] as defined in Section
21 22675 of the Business and Professions Code" that "had at least
22 1,000,000 California users during the preceding 12 months,"
23 §§ 20512(h), 20513-20516, has five main components.

29. First, a requirement that covered platforms “develop and implement procedures for the use of state-of-the-art techniques to identify and remove certain materially deceptive content”⁵ about “candidate[s] for elective office,”⁶ “elections official[s],”⁷ and “elected official[s]”⁸ (the “**Removal Requirement**”). See § 20513.

30. Second, a requirement that covered platforms “develop and implement procedures for the use of state-of-the-art techniques to identify materially deceptive content and for labeling such content” meeting certain conditions (the “**Labeling Requirement**”). See § 20514.

⁵ “Materially deceptive content” means “audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and the output of chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media,” but “does not include any audio or visual media that contains only minor modifications that do not significantly change the perceived contents or meaning of the content,” including “changes to the brightness or contrast of images, removal of background noise in audio, and other minor changes that do not impact the content of the image or audio or visual media.” § 20512(i).

⁶ While AB 2655 does not define “elective office,” “[c]andidate” means any person running for President or Vice President of the United States, any person running for the office of Superintendent of Public Instruction, or any person running for a voter-nominated office as defined in Cal. Elec. Code § 359.5 (see § 20512(c)), which means a “congressional or state elective office for which a candidate may choose to have his or her party preference or lack of party preference indicated upon the ballot” and includes the Governor, Lieutenant Governor, Secretary of State, Controller, Treasurer, Attorney General, Insurance Commissioner, Member of the State Board of Equalization, United States Senator, Member of the United States House of Representatives, State Senator, and Member of the Assembly.

⁷ “Elections official” means (i) the California Secretary of State or (ii) an elections official as defined by Cal. Elec. Code § 320 (§ 20512(g)), which is a (a) “clerk or any person who is charged with the duty of conducting an election,” or (b) “county clerk, city clerk, registrar of voters, or elections supervisor having jurisdiction over elections within any county, city, or district within the state.”

⁸ AB 2655 does not define “elected official.”

1 31. Third, a requirement that covered platforms “provide an
2 easily accessible way for California residents to report to that
3 platform content that should be removed pursuant to Section 20513
4 or labeled pursuant to Section 20514” and “respond to the person
5 who made the report within 36 hours” (the “**Reporting Requirement**”).
6 See § 20515(a).

7 32. Fourth, enforcement provisions, whereby candidates for
8 elective office, elected officials, election officials, the
9 California Attorney General, any California district attorney, and
10 any California city attorney may seek, under certain conditions,
11 “injunctive or other equitable relief against” the covered platform
12 to force it to comply with the Removal Requirement (i.e., to remove
13 particular content), the Labeling Requirement (i.e., to label
14 particular content), or the Reporting Requirement (the “**Enforcement**
15 **Provisions**”). See §§ 20515(b), 20516.

16 33. Fifth, exemptions for certain entities, including
17 broadcasting stations and online newspapers and magazines meeting
18 certain conditions, and certain content, including materially
19 deceptive content that constitutes “satire or parody” (which are
20 terms that the statute does not define). See §§ 20513(d), 20519.

21 **a. The Removal Requirement**

22 34. AB 2655’s Removal Requirement mandates that covered
23 platforms develop and implement procedures that use state-of-the-
24

1 art techniques to identify and remove materially deceptive content
2 if *all* of the following conditions are met, **§ 20513(a)** :

3 a. The content is reported pursuant to Section 20515(a),

4 **§ 20513(a) (1) ;**

5 b. The materially deceptive content is *any* of the following:

6 i. A candidate for elective office portrayed as doing or
7 saying something that the candidate did not do or say
8 and that is reasonably likely to harm the reputation or
9 electoral prospects of a candidate, **§ 20513(a) (2) (A) ;**

10 ii. An elections official portrayed as doing or saying
11 something in connection with the performance of their
12 elections-related duties that the elections official did
13 not do or say and that is reasonably likely to falsely
14 undermine confidence in the outcome of one or more
15 election contests, **§ 20513(a) (2) (B) ;** or

16 iii. An elected official portrayed as doing or saying
17 something that influences an election in California that
18 the elected official did not do or say and that is
19 reasonably likely to falsely undermine confidence in the
20 outcome of one or more election contests,
21 **§ 20513(a) (2) (C) ;**

22 c. The content is posted during the 120 days leading up to an
23 election and through the election day, or – if the content
24 depicts or pertains to elections officials – during the 120

1 leading up to an election, through the election day, and
2 until the 60th day following the election, §§ 20513(a)(3),
3 20513(e); and

4 d. The covered platform knows or acts with reckless disregard
5 for the fact that the content meets Section 20513's
6 requirements, § 20513(a)(4).

7 35. If content "is determined" to meet Section 20513(a)'s
8 requirements, the covered platform must remove the content "upon
9 that determination, but no later than 72 hours after a report is
10 made pursuant to" Section 20515(a). § 20513(b).

11 36. Covered platforms must also identify, using state-of-
12 the-art techniques, and remove, upon discovering or being alerted
13 to the posting or reposting of, any "identical or substantially
14 similar" materially deceptive content that the platform previously
15 removed pursuant to AB 2655, provided that the removal occurs
16 during the time period or periods set forth under Section 20513(e).
17 § 20513(c).

18 **b. The Labeling Requirement**

19 37. AB 2655's Labeling Requirement mandates that covered
20 platforms develop and implement procedures using state-of-the-art
21 techniques to identify materially deceptive content and for
22 labeling such content if all of the following conditions are met,
23 § 20514(a):
24

1 a. The content is reported pursuant to Section 20515(a),
2 **§ 20514(a)(1)**;

3 b. The materially deceptive content is either (i) encompassed
4 by Section 20513(a) but is posted outside Section 20513(e)'s
5 applicable time periods or (ii) appears within an
6 advertisement or election communication⁹ and is not subject
7 to Section 20513, **§ 20514(a)(2)**; and

8 c. The covered platform knows or acts with reckless disregard
9 for the fact that the materially deceptive content meets
10 Section 20514's requirements, **§ 20514(a)(3)**.

11 38. If content "is determined" to meet Section 20514(a)'s
12 requirements, the covered platform must label the content "upon
13 that determination, but no later than 72 hours after a report is
14 made pursuant to" Section 20515(a). **§ 20514(b)**.

15 39. The label required by Section 20514(b) must state: "This
16 [image, audio, or video (depending on the type of content at issue)]
17 has been manipulated and is not authentic." **§ 20514(c)**. The label
18 must also permit users to "click or tap on it for additional
19

20 ⁹ "Election communication" means a general or public communication that is not
21 an "advertisement" and that concerns (i) a candidate for elective office
22 (ii) voting or refraining from voting in an election in California, (iii) the
23 canvass of the vote for an election in California (meaning any election where a
24 "candidate" is on the ballot or where a statewide initiative or statewide
referendum measure is on the ballot), (iv) voting machines, ballots, voting
sites, or other property or equipment related to an election in California, or
(v) proceedings or processes of the electoral college in California.
§§ 20512(e), 20512(f). "Advertisement" means any general or public
communication that a large online platform knows is authorized or paid for with
the purpose of supporting or opposing a candidate for elective office.
§ 20512(a).

1 explanation about the materially deceptive content in an easy-to-
2 understand format.” § 20514(d).

3 40. The Labeling Requirement applies (i) during the period
4 beginning six months before an election in California and through
5 the day of the election; and (ii) if the content depicts or pertains
6 to elections officials, the electoral college process, a voting
7 machine, ballot, voting site, or other equipment related to an
8 election, or the canvass of the vote, during the period beginning
9 six months before an election in California, through the 60th day
10 following the election. § 20514(e).

11 **c. The Reporting Requirement**

12 41. AB 2655’s Reporting Requirement mandates that covered
13 platforms provide an “easily accessible way” for California
14 residents to report to the platform content that should be removed
15 pursuant to Section 20513 or labeled pursuant to Section 20514.
16 § 20515(a).

17 42. The covered platform must respond to the person who made
18 the report within 36 hours of the report, and the response must
19 describe “any action taken or not taken” by the platform with
20 respect to the reported content. *Id.*

21 **d. The Enforcement Provisions**

22 43. AB 2655 provides various methods of enforcement against
23 covered platforms that do not sufficiently comply with the
24 statute’s Removal, Labeling, and Reporting Requirements.

1 44. First, AB 2655 authorizes candidates for elective office,
2 elected officials, and elections officials to seek injunctive or
3 other equitable relief against a covered platform if they make a
4 report pursuant to Section 20515(a) and (i) do not receive a
5 response within 36 hours, (ii) disagree with the platform's
6 response or action taken, or (iii) if the platform does not act
7 within 72 hours. Upon any of those occurrences, AB 2655 authorizes
8 candidates for elective office, elected officials, and elections
9 officials to seek injunctive or other equitable relief against the
10 covered platform to compel (a) the removal of specific content
11 pursuant to Section 20513, (b) the labeling of specific content
12 pursuant to Section 20514, or (c) compliance with the reporting
13 process pursuant to Section 20515(a). There is no action
14 authorized that permits injunctive or equitable relief by any of
15 these parties against covered platforms to compel the platforms to
16 put content back online that was removed improperly or to take down
17 a label of content that was improperly added. **§ 20515 (b) .**

18 45. Second, AB 2655 authorizes the California Attorney
19 General, any California district attorney, and any California city
20 attorney to seek injunctive or other equitable relief against a
21 covered platform to compel (i) the removal of specific content
22 pursuant to Section 20513, (ii) the labeling of specific content
23 pursuant to Section 20514, or (iii) compliance with the reporting
24 process pursuant to Section 20515(a). There is no action

1 authorized that permits injunctive or equitable relief by any of
2 these parties against covered platforms to compel the platforms to
3 put content back online that was removed improperly or to take down
4 a label of content that was improperly added. **§ 20516.**

5 **e. Exemptions**

6 46. AB 2655 exempts certain entities and content from its
7 requirements.

8 47. First, AB 2655 does not apply to regularly published
9 online newspapers, magazines, or other periodicals of general
10 circulation that routinely carry news and commentary of general
11 interest, even if they publish materially deceptive content that a
12 covered platform would be required to remove or label, so long as
13 the publication of the newspaper, magazine, or other periodical
14 contains a "clear disclosure" that the materially deceptive content
15 does not accurately represent any actual event, occurrence,
16 appearance, speech, or expressive conduct. **§ 20519(a).**

17 48. Second, AB 2655 does not apply to broadcasting stations
18 that broadcast prohibited materially deceptive content as part of
19 a "bona fide newscast, news interview, news documentary, commentary
20 of general interest, or on-the-spot coverage of bona fide news
21 events," so long as the broadcast "clearly acknowledges," through
22 content or a disclosure and in a manner that can be "easily heard
23 or read by the average listener or viewer," that the materially
24 deceptive content does not accurately represent any actual event,

1 occurrence, appearance, speech, or expressive conduct.

2 **§ 20519(b)(1)**.

3 49. Third, AB 2655 does not apply to broadcasting stations
4 that are paid to broadcast materially deceptive content if (i) the
5 broadcasting station can show that it has "prohibition and
6 disclaimer requirements that are consistent" with those set forth
7 in the statute and has provided those requirements to each person
8 or entity that purchased the advertisement, or (ii) federal law
9 requires that the broadcasting station air advertisements from
10 legally qualified candidates or prohibits the broadcasting station
11 from censoring or altering the message. **§ 20519(b)(2)**.

12 50. Fourth, AB 2655 does not apply to materially deceptive
13 content that constitutes "satire or parody." **§ 20519(c)**.

14 51. Finally, AB 2655's Removal Requirement does not apply to
15 a candidate for elective office who, during the time period set
16 forth in Section 20513(e), "portrays themselves" as doing or saying
17 something that the candidate did not do or say, if the digital
18 content includes a disclosure stating: "This [image, audio, or
19 video (depending on the type of content at issue)] has been
20 manipulated." **§ 20513(d)**.

21 a. For visual media, the text of the disclosure must be in a
22 size that is "easily readable by the average viewer and no
23 smaller than the largest font size of other text appearing
24 in the visual media." If the visual media includes no other

1 text, the disclosure must be "in a size that is easily
 2 readable by the average viewer." For visual media that is
 3 video, the disclosure shall appear for the duration of the
 4 video. **§ 20513(d)(2)(A)**.

5 b. If the media consists of audio only, the disclosure must be
 6 read in a "clearly spoken manner and in a pitch that can be
 7 easily heard by the average listener, at the beginning of
 8 the audio, at the end of the audio, and, if the audio is
 9 greater than two minutes in length, interspersed within the
 10 audio at intervals of not greater than two minutes each."

11 **§ 20513(d)(2)(B)**.

12 **II. AB 2655 Imposes Content-Based Restrictions on Protected**
 13 **Political Speech**

14 52. The legislative history of AB 2655 is riddled with
 15 numerous references to the First Amendment problems raised by the
 16 statute. As the legislative history makes clear, by explicitly
 17 targeting derogatory political speech about candidates, AB 2655
 18 imposes content-based speech restrictions that, under our
 19 Constitution and precedents, must be given the "broadest
 20 protection" to maintain a free-flowing marketplace of ideas for
 21 the "bringing about of political and social changes desired by the
 22 people." See *McIntyre*, 514 U.S. at 346. For instance:

23 53. The Assembly Committee on Judiciary's April 22, 2024
 24 analysis acknowledges that

[AB 2655] would **interfere with both the expression and reception of information based upon its content.** Moreover, **not only does this bill single out particular content, the content relates to political candidates and elections.** This is potentially problematic because the First Amendment affords the **"broadest protection"** to the "discussion of public issues" and "political expression in order to assure the unfettered interchange of ideas for the bringing about of political and social changes desired by the people." (*McIntyre v Ohio Election Commission* (1997) 514 U.S. 334.) It is difficult to imagine any content more related to "political expression" and "discussion of public issues" than content about candidates and elections. **The fact that the bill restricts speech that is "materially deceptive" or "false" does not matter,** for the U.S. Supreme Court has been unequivocal that the First Amendment protects even "false" speech. **The remedy for false speech is more true speech, and false speech tends to call forth true speech.** (*United States v Alvarez* (2012) 567 U.S. 709.)

Ex. 4 (Assemb. Standing Comm. on Judiciary, Analysis of Assemb. Bill No. 2655, 2023-2024 Reg. Sess. (Cal. Apr. 22, 2024)) at 7.

54. The Senate Judiciary Committee's June 28, 2024 analysis states that **"[l]aws that burden political speech are subject to strict scrutiny . . . California courts have been clear that political expression in the context of campaigns of any manner should be given wide latitude[.]"** Ex. 3 at 14 (citing *Citizens United v. Fed. Election Comm'n*, 558 U.S. 310, 340 (2010); *Beilenson v. Superior Ct.*, 44 Cal. App. 4th 944, 954-55 (1996)).

55. The Assembly Committee on Judiciary's April 22, 2024 analysis recognizes that **"[i]n reviewing the law, the Court would apply strict scrutiny."** Ex. 4 at 8.

56. California State Assembly member Rebecca Bauer-Kahan, who supported AB 2655, stated, **"I think we all agree that strict**

1 **scrutiny would be applied."** Ex. 5 (*Defending Democracy from*
 2 *Deepfake Deception Act of 2024: Hearing on AB 2655 Before the*
 3 *Assemb. Standing Comm. on Judiciary, 2023-2024 Reg. Sess. (Cal.*
 4 *Apr. 23, 2024)*) at 6 (statements of Rebecca Bauer-Kahan, Assemb.
 5 Member).¹⁰

6 57. The American Civil Liberties Union, which opposed AB
 7 2655, explained that the

8 **"novelty of deepfake technology and the speed with which**
 9 **it is improving" do not justify relaxing the stringent**
 10 **protections afforded to political speech by the First**
 11 **Amendment.** The Supreme Court has held that "whatever the
 12 challenges of applying the Constitution to ever advancing
 13 technology, **'the basic principles of freedom of speech**
 14 **and the press, like the First Amendment's command, do**
 15 **not vary' when a new and different medium for**
 16 **communication appears."** The law has long made clear that
 17 the First Amendment was intended to create a **wide berth**
 18 **for political speech** because it is the core of our
 19 democracy. The First Amendment provides robust
 20 protection for speech of all kinds. Speech that is false,
 21 confusing, or which presents content that some find
 22 abhorrent, nevertheless maintains its constitutional
 23 protections as a driver of free discourse. This remains
 24 so no matter what the technology used to speak.
Unfortunately, the provisions of AB 2655 as currently
drafted threaten to intrude on those rights and deter
that vital speech.

Ex. 3 at 18-19.

¹⁰ Available at
<https://digitaldemocracy.calmatters.org/hearings/257837?t=255&f=afb99536b82e1a34379ebbfd23fe84b1> (4:37-4:40) (last visited Nov. 14, 2024). All exhibit transcripts, which were downloaded directly from the websites, are auto-generated, uncertified, and may contain errors. To that end, all quotations herein are transcribed directly from the videos themselves.

1 **III. AB 2655 Will Result in Censorship of Substantial Amounts of**
2 **Valuable Political Speech**

3 58. Whether content is prohibited under AB 2655 hinges on
4 various undefined terms that render it impossible for covered
5 platforms to comply with the statute in a precise manner. Moreover,
6 because the Enforcement Provisions provide for causes of action
7 seeking to require the covered platforms to remove or label
8 “materially deceptive content” covered by the statute, but do not
9 provide for any consequences for improperly removing or labeling
10 content that should not have been removed or labeled, the covered
11 platforms are incentivized under the enforcement regime to err
12 significantly on the side of censorship to avoid the substantial
13 costs associated with defending lawsuits under the statute. And,
14 as AB 2655’s legislative history makes clear, this will result in
15 substantial censorship of content that lies at the heart of the
16 protections provided by the First Amendment – including important
17 commentary that invites vital discussion about election officials
18 and candidates.

19 59. The April 8, 2024 analysis of the Assembly Committee on
20 Elections aptly describes the difficulties that covered platforms
21 will encounter in attempting to comply with AB 2655:

22 [I]n order to determine whether it must block content
23 that *portrays a candidate for election as doing or saying*
24 *something that the candidate did not do or say*,¹¹ the
platform would need to know not only that the person
portrayed in the content was a candidate for office, but

¹¹ Emphasis in original.

also the date (or dates) of the election when the candidate will appear on the ballot. Similarly, it would need to determine whether the candidate had actually said or done the thing that the candidate is portrayed as doing. While some of that information will be widely available and well known in some cases (e.g., the identity of major party candidates for President of the United States in presidential general elections and the dates of federal elections), it will be more arcane in other situations. Given the number of elections (including standalone local and special elections) and candidates (including write-in candidates and candidates for local elections in smaller jurisdictions) in California at any given time, **making the determinations at scale about which content must be blocked or labeled likely will be considerably more challenging than making those determinations on a case-by-case basis in a court of law.**

Ex. 6 (Assemb. Standing Comm. on Elections, Analysis of Assemb. Bill No. 2655, 2023–2024 Reg. Sess. (Cal. Apr. 8, 2024)) at 8.

60. The statute's compressed timeframes for making these determinations – covered platforms must respond to requests to remove content pursuant to the statute “within 36 hours, describing any action taken or not taken” with respect to the content, § 20515(a), and take action to remove any such content “no later than 72 hours after a report is made,” § 20513(b) – only exacerbate these problems. If these timeframes are not met, an enforcement action may be filed against the covered platform. See §§ 20515(b), 20516.

61. Tracy Rosenberg of Oakland Privacy, which opposed AB 2655, similarly recognized that “**technology platform[s] can[not] be expected to know everything that every candidate running for office [has said] . . . So basically we’re using imprecise measures**

1 **to power a potentially broad censorship regime of blocking content.**
 2 **And we really can't support that even under the guise of defending**
 3 **democracy."** Ex. 7 (*Defending Democracy from Deepfake Deception*
 4 *Act of 2024: Hearing on AB 2655 Before the Assemb. Standing Comm.*
 5 *on Elections*, 2023-2024 Reg. Sess. (Cal. Apr. 10, 2024)) at 6
 6 (statements of Tracy Rosenberg, Oakland Privacy).¹² At a hearing
 7 in front of the Senate Committee on Judiciary, Rosenberg added that
 8 **"[t]his is not what people want."** Ex. 8 (*Defending Democracy from*
 9 *Deepfake Deception Act of 2024: Hearing on AB 2655 Before the S.*
 10 *Standing Comm. on Judiciary*, 2023-2024 Reg. Sess. (Cal. July 2,
 11 2024)) at 4 (statements of Tracy Rosenberg, Oakland Privacy).¹³

12 62. Difficult questions about the applicability of the
 13 statute to any given political advertisement or video will be
 14 commonplace and will put covered platforms in a bind; they can
 15 either remove or label any content raising close calls (and avoid
 16 entirely the risk of liability) or subject themselves to a high
 17 likelihood of costly litigation.

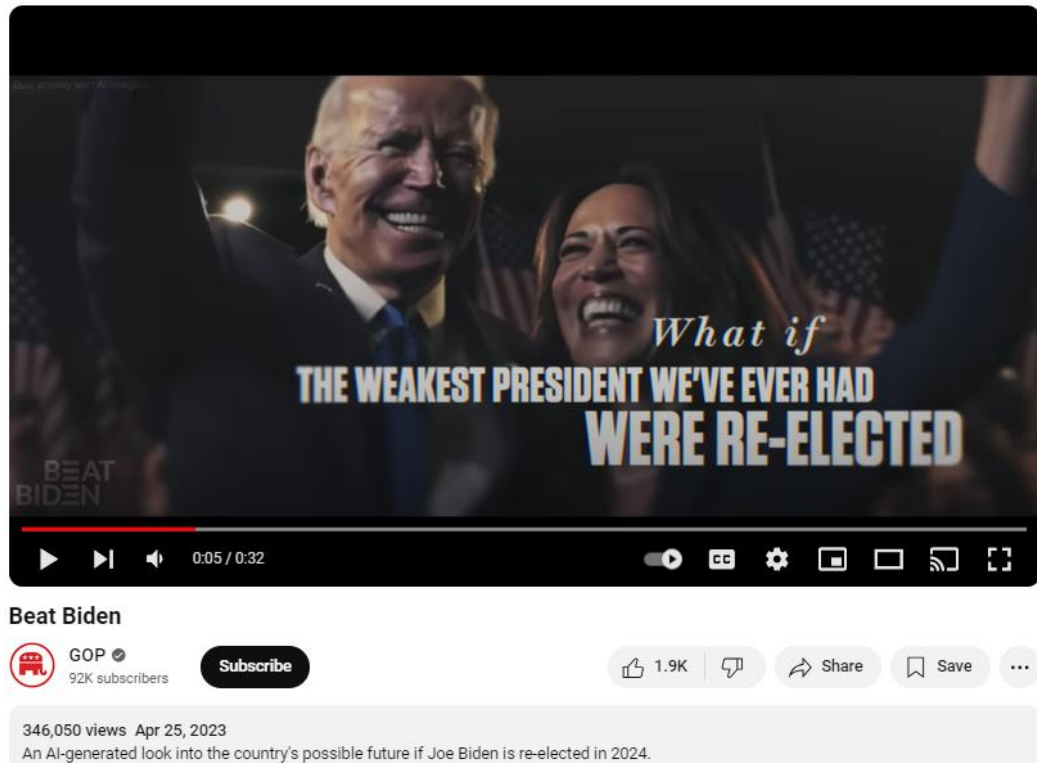
18 63. For instance, on April 25, 2023, the official Republican
 19 National Committee YouTube channel posted a video titled "Beat
 20 Biden" that, using artificial intelligence, imagined various
 21 scenarios that would occur during a second presidential term under

22 ¹² Available at
 23 <https://digitaldemocracy.calmatters.org/hearings/257736?t=1986&f=da025f00cb70d1ea6196340ca76df63e> (33:23-34:12) (last visited Nov. 14, 2024).

24 ¹³ Available at
<https://digitaldemocracy.calmatters.org/hearings/258109?t=763&f=7421e586be4213e768ac887bce75f630> (12:54) (last visited Nov. 14, 2024).

Joe Biden, including that “international tensions [will] escalate,” “financial systems [will] crumble,” and “crime [will] worsen[.]” Ex. 2. As shown below in Figure 1, the video’s description states that it is “[a]n AI-generated look into the country’s possible future if Joe Biden is re-elected in 2024.”

FIGURE 1



64. Does this video portray President Biden “doing or saying something that” he “did not do or say,” and would it have been “reasonably likely” that the video would have “harm[ed] [his] reputation or electoral prospects?” Perhaps not, but this video was cited in AB 2655’s legislative history as an example of how “generative AI can spread misinformation regarding elections with ease,” see Ex. 3 at 7, 9, seemingly indicating that, at least some

1 of the drafters think it would be prohibited under the statute.
2 Given that the video asks "what if the weakest president we've ever
3 had were re-elected," would the video fall within Section
4 20519(c)'s exemption for satire or parody? That is also unclear.
5 Adding to the confusion, moreover, is that the video's caption
6 clearly states that the video was "AI-generated," but this would
7 not bring the video within Section 20513(d)'s safe harbor because
8 it was posted by someone other than President Biden. See § 20513(d)
9 ("[T]his section does not apply to a candidate for elective office
10 who . . . **portrays herself** as doing or saying something that the
11 candidate did not do or say . . . "). Faced with this lack of
12 clarity, and while having to make this type of determination at
13 mass-scale, covered platforms would have no choice but to remove
14 the video or potentially face enforcement actions brought by highly
15 motivated political opponents or government officials.

16 65. Another example further demonstrates AB 2655's
17 unintelligibility. In March 2023, an X user named Eliot Higgins
18 (@EliotHiggins) used artificial intelligence to create a photo
19 depicting Donald Trump being forcefully arrested. Ex. 1; see
20 Figure 2, below. The same questions arise. Do these photos portray
21 Donald Trump "doing or saying something that" he "did not do or
22 say," and would it be "reasonably likely" that the photos would
23 "harm [his] reputation or electoral prospects?" Would these photos
24 be exempted as satire or parody under Section 20519(c)? As long

as colorable arguments can be made that this type of political commentary is covered by the statute, covered platforms will be faced with the choice of removing and/or labeling such content (which would ensure no liability for them) or facing costly enforcement actions.

FIGURE 2



66. To take another example, on August 29, 2024, the X user Kamala HQ (@KamalaHQ) posted a five-second video on X where Vice Presidential candidate JD Vance says, "Democrats want to attack Republicans as being anti-union and sometimes the shoe fits."¹⁴ The clip cuts out right before Vance says "but not me, and not Donald

¹⁴ Ex. 9 (Kamala HQ (@KamalaHQ), X (Aug. 29, 2024, 12:57 PM), <https://x.com/KamalaHQ/status/1829201653175636390> (last visited Nov. 14, 2024)).

1 Trump.”¹⁵ How would the statute treat this edited snippet, which
2 arguably misleadingly changes the **meaning** of what JD Vance actually
3 said? AB 2655 defines “materially deceptive content” as “audio or
4 visual media that is digitally created **or modified** . . . such that
5 it would falsely appear to a reasonable person to be an authentic
6 record of the content depicted in the media.” § 20512(i)(1). In
7 the context of highly contested elections, candidates and
8 government officials (such as Defendants) would be incentivized to
9 issue take down requests for videos, like this one, that have even
10 arguably been modified in ways that change their meaning and
11 arguably give a misleading impression of what was actually said.
12 The results would be calamitous. To avoid liability, covered
13 platforms will be incentivized to remove and/or label such content
14 pursuant to the statute. If they fail to do so, they will likely
15 face costly enforcement actions.

16 67. Finally, AB 2655 purports to exempt “[m]aterially
17 deceptive content that constitutes satire or parody,” § 20519(c),
18 but it does not define “satire or parody.” When faced with
19 arguments about whether otherwise “materially deceptive content”
20 encompassed by the statute is “satire” or “parody,” AB 2655
21 incentivizes covered platforms to remove and/or label such content
22 whenever there is a debate about that highly contentious subject.

23 _____
24 ¹⁵ See the full video at Ex. 10 (The International Association of Fire Fighters,
57th IAFF Convention: Sen. JD Vance, YouTube (Aug. 29, 2024),
<https://www.youtube.com/watch?v=EGKTo5j3gl0&t=1081s> (last visited Nov. 14,
2024)).

1 This is because, under the Enforcement Provisions, removal and/or
2 labeling of flagged content results in complete immunity for the
3 covered platforms, while refusing to do so opens them up to
4 potential costly litigation.

5 68. Consider the video posted by Christopher Kohls, a content
6 creator who goes by the name Mr. Reagan, titled *Kamala Harris Ad*
7 *PARODY*, that was reposted on X by Elon Musk.¹⁶ The video uses AI
8 to create an "advertisement" by Vice President Harris that has her
9 saying things that she would never actually say. While some would
10 reasonably consider the video to be satire or parody – including
11 because, in the video, "Harris" states that she is a "diversity
12 hire," who "may not know the first thing about running the country"
13 and is a "deep state puppet" – public statements made by Governor
14 Newsom indicate that he believes that the statute would require
15 the video to be removed from any covered platform. See Ex. 13
16 (Gavin Newsom (@GavinNewsom), X (Sept. 17, 2024, 7:41 PM),
17 <https://x.com/GavinNewsom/status/1836188721663873324> (last
18 visited Nov. 14, 2024)) (stating that Mr Reagan's *Kamala Harris Ad*
19 *PARODY* video "should be illegal" and declaring, the same day that
20 AB 2655 was passed, that he "just signed a bill to make this illegal
21 in the state of California"). Under AB 2655, for covered platforms
22

23 ¹⁶ See Ex. 11 (Mr Reagan, *Kamala Harris Ad PARODY*, YouTube (July 26, 2024),
24 <https://www.youtube.com/watch?v=sVSpeqNnoWM> (last visited Nov. 14, 2024)); see
also Ex. 12 (Elon Musk (@elonmusk), X (July 26, 2024, 7:11 PM),
<https://x.com/elonmusk/status/1816974609637417112> (last visited Nov. 14,
2024)).

1 to protect such speech, they will have to pay dearly by defending
2 their content-moderation decisions in court. And if they remove
3 such content, they will have no costs at all.

4 69. This combination of AB 2655's unintelligible
5 requirements and draconian and one-sided Enforcement Provisions –
6 which protect removal of content from any liability and impose
7 enforcement costs only on decisions not to remove content – will
8 lead to censorship at the direction of the State. Liability
9 regimes, set up by the State, that have a “tendency to inhibit
10 constitutionally protected expression” cannot survive First
11 Amendment scrutiny. *Smith v. California*, 361 U.S. 147, 155 (1959)
12 (striking down, on First Amendment grounds, city ordinance
13 providing for strict liability for possession of books later judged
14 to be obscene).

15 70. AB 2655's legislative history openly acknowledges the
16 serious First Amendment problems raised by the statute's incentive
17 structure and enforcement regime. For instance:

18 71. The Assembly Committee on Judiciary's April 22, 2024
19 analysis acknowledges that, “[c]onfronted with such a restricted
20 timeline and the threat of a civil action . . . **platforms will**
21 **'remove significantly more content, including content that has**
22 **accurate election information and content that is not materially**
23 **deceptive.'**” Ex. 4 at 12.
24

72. The analysis also recognizes that **"with no sure means to determine what is 'materially deceptive,' the platforms will err on the side of blocking content, thus burdening more speech than is necessary."** *Id.* at 8.

73. Jose Torres Casillas of TechNet, which opposed AB 2655, explained that AB 2655

[R]equires online platforms to make determinations about truth and falsity in an impossible way. Instances where content or information is clearly true or clearly false are not [the] norm. Far more often, content falls into a middle ground where it requires time and a fact-intensive investigation to determine whether something is true or false. Investigative journalists have challenges with fact checking even the most high profile races or candidates. It is difficult enough for a platform to know whether something is false as it relates to a presidential candidate or a high profile federal race, and this is simply impossible for races lower down on the ticket. **A platform cannot accurately adjudicate reports on those types of content and will instead resort to over removing information in order to avoid liability and the penalties in this bill. Removing information that is only suspected of being false is clearly not a good outcome.**

Ex. 5 at 5 (statements of Jose Torres Casillas, TechNet).¹⁷

74. Khara Boender of the Computer Communications Industry Association (CCIA), which also opposed AB 2655, similarly explained that the content-moderation "tools that are currently available [to covered platforms] are not always reliable or accurate," and

Because covered platforms are not privy to the intent and context for which a piece of content is used, they could **inadvertently over block or over label content.**

¹⁷ Available at <https://digitaldemocracy.calmatters.org/hearings/257837?t=145&f=afb99536b82e1a34379ebbfd23fe84b1> (2:39-3:38) (last visited Nov. 14, 2024).

This could result in user frustration and suppression of political speech. Political speech was at the core of why our First Amendment was established, and it is critical to maintain those protections. ***Responsibility for labeling AI generated election content and liability for the deceptive content should rest with the entity that puts forth such material,*** the one that is most aware of the intent and context for which the content was created and shared. . . . And while the bill exempts satire and parody, it is unclear who gets to decide what constitutes those uses. ***Faced with individual users seeking injunctive relief merely if they disagree with a covered platform's decision regarding reported content, a service may choose to prohibit all digitally altered content, cutting off many valuable and helpful uses.***

Id. at 4–5 (statements of Khara Boender, CCIA).¹⁸

75. Boender explained that AB 2655 will have an effect similar to that of the takedown regime under the Digital Millennium Copyright Act (DMCA), which, like AB 2655, provides immunity from liability if material is taken down but potential liability if it is not. See 17 U.S.C. § 512(c)(1). As Boender correctly pointed out, AB 2655 will “**result in platforms being required to block content almost constantly in order to ensure compliance,**” which has been the outcome under the DMCA, where platforms “err in taking down the content lest they face[] liability.” Ex. 14 (*Defending Democracy from Deepfake Deception Act of 2024: Hearing on AB 2655 Before the S. Standing Comm. on Elections and Constitutional Amends.*, 2023–2024 Reg. Sess. (Cal. June 18, 2024)) at 5 (statements of Khara Boender, CCIA);¹⁹ see also Ex. 15 (Wendy

¹⁸ Available at <https://digitaldemocracy.calmatters.org/hearings/257837?t=27&f=afb99536b82e1a34379ebbfd23fe84b1> (0:49–2:09) (last visited Nov. 14, 2024).

¹⁹ Available at

1 Seltzer, *Free Speech Unmoored in Copyright's Safe Harbor: Chilling*
 2 *Effects of the DMCA on the First Amendment*, 24 Harv. J.L. & Tech.
 3 171 (2010)) (asserting that the DCMA encourages internet service
 4 providers to respond to copyright complaints by removing content
 5 to ensure immunity from liability, leading to the censorship of
 6 protected speech).

7 76. California Assembly Member Bill Essayli, who opposed the
 8 bill, recognized that ABB 2655's requirements are "a very sticky
 9 thing with the First Amendment and also with asking private
 10 companies to be the enforcer," and expressed that a better
 11 alternative is "***the Twitter model where they use the community to***
 12 ***sort of regulate information on there. . . . where it's the public,***
 13 ***it's the crowd sourcing, is kind of doing the moderating,*** rather
 14 ***than "making an individual, company, or person the arbiter of***
 15 ***what's disinformation."*** Ex. 7 at 7-8 (statements of Bill Essayli,
 16 Assemb. Member).²⁰

23 [https://digitaldemocracy.calmatters.org/hearings/258097?t=87&f=213a711036e0125](https://digitaldemocracy.calmatters.org/hearings/258097?t=87&f=213a711036e0125a7084c8b0dee7c131)
 24 [a7084c8b0dee7c131](https://digitaldemocracy.calmatters.org/hearings/258097?t=87&f=213a711036e0125a7084c8b0dee7c131) (1:38-2:15) (last visited Nov. 14, 2024).

²⁰ Available at
[https://digitaldemocracy.calmatters.org/hearings/257736?t=2285&f=da025f00cb70d](https://digitaldemocracy.calmatters.org/hearings/257736?t=2285&f=da025f00cb70d1ea6196340ca76df63e)
[1ea6196340ca76df63e](https://digitaldemocracy.calmatters.org/hearings/257736?t=2285&f=da025f00cb70d1ea6196340ca76df63e) (38:10-38:51) (last visited Nov. 14, 2024).

1 **IV. AB 2655 Impermissibly Substitutes the Government's Judgment**
2 **About Content Moderation for That of the Covered Platforms**

3 77. X already has its own policy for regulating "synthetic"
4 or "manipulated media" on its platform. Under X's Synthetic and
5 Manipulated Media Policy, users "may not share synthetic,
6 manipulated, or out-of-context media that may deceive or confuse
7 people and lead to harm ('misleading media')." In addition, under
8 the policy X "may label posts containing misleading media to help
9 people understand their authenticity and to provide additional
10 context." Ex. 16 (*Synthetic and manipulated media policy*, X,
11 <https://help.x.com/en/rules-and-policies/manipulated-media> (last
12 visited Nov. 14, 2024)) at 3.

13 78. Under X's policy – which is publicly available to all
14 users of the platform as well as to the public generally – X uses
15 the following criteria when considering removal and/or labeling of
16 posts and media:

- 17 • 1. Is the content significantly and deceptively altered,
18 manipulated, or fabricated?
- 19 • 2. Is the content shared in a deceptive manner or with false
20 context?
- 21 • 3. Is the content likely to result in widespread confusion
22 on public issues, impact public safety, or cause serious
23 harm?

24 *See id.*

1 79. In addition, X's policy also makes clear that the
2 following are "generally not in violation of this policy":

- 3 • **Memes or satire**, provided these do not cause
4 significant confusion about the authenticity of the
5 media.
- 6 • **Animations, illustrations, and cartoons**, provided
7 these do not cause significant confusion about the
8 authenticity of the media.
- 9 • **Commentary, reviews, opinion, and/or reactions**.
10 Sharing media with edits that only add commentary,
11 reviews, opinions, or reactions allows for further
12 debate and discourse relating to various issues are
13 not in violation of this policy.
- 14 • **Counterspeech**. We allow for direct responses to
15 misleading information which seek to undermine its
16 impact by correcting the record, amplifying credible
17 information, and educating the wider community about
18 the prevalence and dynamics of misleading information.

19 See *id.* at 6.

20 80. Other covered platforms (e.g., Meta, YouTube, TikTok,
21 and Snapchat) all have their own policies designed to address
22 false, misleading, and/or manipulated media. See Ex. 17 (*How to*
23 *identify AI content on Meta products*, Meta,
24 <https://www.meta.com/help/artificial-intelligence/how-ai->

generated-content-is-identified-and-labeled-on-meta/ (last visited Nov. 14, 2024)) at 2 ("Meta requires an AI label when content has photorealistic video or realistic-sounding audio that was digitally created, modified or altered, including with AI."); Ex. 18 (*Disclosing use of altered or synthetic content*, YouTube, <https://support.google.com/youtube/answer/14328491> (last visited Nov. 14, 2024)) at 1 ("To help keep viewers informed about the content they're viewing, we require creators to disclose content that is meaningfully altered or synthetically generated when it seems realistic."); Ex. 19 (*About AI-generated content*, TikTok, <https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content> (last visited Nov. 14, 2024)) at 5 ("We also require creators to label all AI-generated content where it contains realistic images, audio, and video, as explained in our Community Guidelines."); Ex. 20 (*Generative AI on Snapchat*, Snapchat, <https://help.snapchat.com/hc/en-us/articles/25494876770580-Generative-AI-on-Snapchat> (last visited Nov. 14, 2024)) at 1 ("We may indicate that a feature in Snapchat is powered by generative AI in a number of ways . . . When you see these contextual symbols or other indicators in Snapchat, you should know that you are . . . viewing content that has been produced using AI and does not depict real world scenarios.").

81. Each platform takes a different approach to these content-moderation decisions, as is the right of each platform

1 under the First Amendment. See *Moody*, 144 S. Ct. at 2394, 2401,
2 2403, 2405, 2409.

3 82. AB 2655 impermissibly substitutes the State's content-
4 moderation policies in this important area for those of the covered
5 platforms' and impermissibly imposes liability on the covered
6 platforms for noncompliance with the State's preferred content-
7 moderation policies. This violates the First Amendment.

8 83. X also currently has a program called "Community Notes"
9 that allow users to flag, among other things, content that they
10 believe needs context, is "materially deceptive" and otherwise
11 covered by the statute, or has been digitally altered. Users are
12 free to provide additional context or information about the content
13 that will appear with the content if enough of the community's
14 "contributors," who otherwise hold diverse points of view, deem
15 the additional commentary to be helpful. And, in recognition of
16 the fast-paced nature of social media, X has accelerated Community
17 Notes and now indicates "Lightning Notes," which start appearing
18 on posts within an hour of being proposed, or within an hour of
19 the post itself going live.

20 84. The State has never explained why X's Synthetic and
21 Manipulated Media Policy, coupled with its "Community Notes"
22 program, are insufficient to address the "materially deceptive
23 content" targeted by AB 2655. In fact, they work very well.
24

1 85. Nor has the State explained why the policies of other
2 covered platforms, coupled with counterspeech from other users of
3 the platforms, are insufficient to address the “materially
4 deceptive content” targeted by AB 2655 in a less speech-restrictive
5 manner.

6 **V. AB 2839 & The Kohls Action**

7 86. On September 17, 2024, the same day Governor Newsom
8 signed AB 2655 into law, he also signed into law AB 2839 (codified
9 at Cal. Elec. Code § 20012), which institutes largely the same
10 requirements as AB 2655 but frames them in terms of potential
11 liability for content creators, rather than for platforms.

12 87. For instance, like AB 2655, AB 2839 prohibits “materially
13 deceptive content” (defined nearly identically across the statutes)
14 that is (i) a “candidate for any federal, state, or local elected
15 office in California portrayed as doing or saying something that
16 the candidate did not do or say if the content is reasonably likely
17 to harm the reputation or electoral prospects of a candidate,”
18 § 20012(b)(1)(A) (*compare with* § 20513(a)(2)(A)); (ii) an
19 “elections official portrayed as doing or saying something in
20 connection with an election in California that the elections
21 official did not do or say if the content is reasonably likely to
22 falsely undermine confidence in the outcome of one or more election
23 contests,” § 20012(b)(1)(B) (*compare with* § 20513(a)(2)(B)); or
24 (iii) an “elected official portrayed as doing or saying something

1 in connection with an election in California that the elected
2 official did not do or say if the content is reasonably likely to
3 harm the reputation or electoral prospects of a candidate or is
4 reasonably likely to falsely undermine confidence in the outcome
5 of one or more election contests," § 20012(b)(1)(C) (*compare with*
6 § 20513(a)(2)(C)).

7 88. As does AB 2655, AB 2839 institutes a *mens rea*
8 requirement. *Compare* § 20012(b)(1) (limiting prohibitions to those
9 that, "with malice, knowingly" violate § 20012(b)) with §§
10 20513(a)(4), 20514(a)(3) (limiting Removal and Labeling
11 Requirements to those that "know[] or act[] with reckless
12 disregard").

13 89. On September 17, 2024, Christopher Kohls, an individual
14 who creates digital content about political figures and who owns
15 the screen name "Mr Reagan" on YouTube, *see supra* ¶ 68, moved for
16 a preliminary injunction in the United States District Court for
17 the Eastern District of California to enjoin the enforcement of AB
18 2839, because it violated (i) the First Amendment of the United
19 States Constitution and Article I, Section 2, of the California
20 Constitution (both facially and as-applied) and (ii) the Fourteenth
21 Amendment of the United States Constitution for vagueness.

22 90. On October 2, 2024, the Honorable John A. Mendez granted
23 the motion, finding that Kohls was likely to succeed in showing
24 that AB 2839 facially violates the First Amendment and Article I,

1 Section 2, of the California Constitution, which is at least as
2 protective, because AB 2839 is a content-based speech restriction
3 that triggers and fails strict scrutiny. *Kohls*, 2024 WL 4374134,
4 at *3-6.

5 91. In *Kohls*, the Court held that AB 2839 triggered
6 constitutional review under strict scrutiny because it
7 “specifically targets speech within political or electoral content
8 pertaining to candidates, electoral officials, and other election
9 communication, making it a content-based regulation that seeks to
10 limit public discourse.” *Id.* at *4.

11 92. The Court held that AB 2839 failed strict scrutiny
12 because it was not the “least restrictive means available for
13 advancing [its] interest,” *id.* (quoting *NetChoice, LLC v. Bonta*,
14 113 F.4th 1101, 1121 (9th Cir. 2024)), since “[o]ther statutory
15 causes of action such as privacy torts, copyright infringement, or
16 defamation already provide recourse to public figures or private
17 individuals whose reputations may be afflicted by artificially
18 altered depictions peddled by satirists or opportunists on the
19 internet,” *id.* at *5.

20 93. The Court also rejected the arguments of defendants
21 Robert Bonta and Shirley Weber that AB 2839 only restricts
22 unprotected defamatory and/or false speech. *See id.* at *3-4. The
23 Court explained that AB 2839 “does not use the word ‘defamation’
24 and by its own definition, extends beyond the legal standard for

1 defamation to include any false or materially deceptive content
2 that is 'reasonably likely' to harm the 'reputation **or** electoral
3 prospects of a candidate,'" *id.* at *3 (quoting § 20012(b))
4 (emphasis in original), and "does much more than punish potential
5 defamatory statements" because it "does not require actual harm
6 and sanctions any digitally manipulated content that is 'reasonably
7 likely' to 'harm' the amorphous 'electoral prospects' of a
8 candidate or elected official," *id.* (quoting §§ 20012(b)(1)(A),
9 (C)).

10 94. The Court further explained that AB 2839 did not restrict
11 speech that was otherwise unprotected as "lies that involve 'some
12 . . . legally cognizable harm'" under *United States v. Alvarez*,
13 567 U.S. 709 (2012), and that AB 2839 imposed "civil penalties for
14 criticisms on the government" that "have no place in our system of
15 governance." *Kohls*, 2024 WL 4374134, at *4.

16 95. All of these arguments as to why AB 2839 fails to satisfy
17 First Amendment scrutiny apply equally to AB 2655.

18 **FIRST CAUSE OF ACTION**

19 **(Declaratory Relief and Preliminary and Permanent Injunctive**
20 **Relief for Violations of the First Amendment of the United States**
Constitution (42 U.S.C. § 1983) and Article I, Section 2, of the
California Constitution – Facial and As-Applied)

21 96. X Corp. realleges and incorporates herein by reference
22 each and every allegation set forth above.

23 97. AB 2655 violates the First Amendment of the United States
24 Constitution and Article I, Section 2, of the California

1 Constitution by forcing covered platforms like X, under threat of
2 injunctive and other equitable enforcement, to remove and alter
3 certain constitutionally protected election-related content of
4 which the State of California disapproves, and to create a
5 reporting procedure to facilitate such removal and alteration.²¹

6 98. First, AB 2655 imposes a prior restraint on speech, which
7 is the "most serious and the least tolerable infringement on First
8 Amendment rights," *Stuart*, 427 U.S. at 559, and does so as to
9 speech concerning "public issues and debate on the qualifications
10 of candidates," to which the "First Amendment affords the **broadest**
11 **protection**" to protect the "unfettered interchange of ideas for
12 the bringing about of political and social changes desired by the
13 people," *McIntyre*, 514 U.S. at 346.

14 99. AB 2655 imposes a prior restraint on speech because
15 Sections 20515(b) and 20516 provide expedited causes of action
16 under Section 35 of the California Code of Civil Procedure through
17 which political speech will be enjoined before there occurs
18 a "final judicial determination" that the "speech is unprotected."
19 *Isaksen*, 2005 WL 8176605, at *3 (citing *Vance*, 445 U.S. 308)

21 ²¹ AB 2655 violates Article I, Section 2, of the California Constitution for all
22 of the same reasons that it violates the First Amendment of the United States
23 Constitution. See, e.g., *Kohls*, 2024 WL 4374134, at *6 ("Under current case
24 law, the California state right to freedom of speech is at least as protective
as its federal counterpart."); *City of Montebello v. Vasquez*, 1 Cal. 5th 409,
421 n.11 (2016) ("[T]he California liberty of speech clause is broader and more
protective than the free speech clause of the First Amendment."); *Delano Farms*
Co. v. California Table Grape Com., 4 Cal. 5th 1204, 1221 (2018) ("[O]ur case
law interpreting California's free speech clause has given respectful
consideration to First Amendment case law for its persuasive value.").

(denying motion for preliminary injunction as to already published speech because it would have constituted a prior restraint). Even if a plaintiff demonstrates “through clear and convincing evidence” that the speech meets the requirements of the statute, that showing **does not** amount to proof that the speech is constitutionally unprotected. See *Kohls*, 2024 WL 4374134, at *3-4; see also *Garcia*, 786 F.3d at 747 (forcing Google through “takedown order” to remove content previously published on YouTube before a final determination that the content was unprotected amounted to a “classic prior restraint on speech”); *Kelley*, 2023 WL 2347442, at *9 (citing *Alexander*, 509 U.S. at 550; *Garcia*, 786 F.3d at 746-47) (prior restraints “refer either to injunctions that restrict future speech or require takedowns of currently-published speech”); *SolarPark Korea Co.*, 2023 WL 4983159, at *11 (same). AB 2655 cannot overcome the “historical and heavy presumption against such restraints.” *Garcia*, 786 F.3d at 747.

100. In addition, AB 2655 imposes a prior restraint on speech because (i) nothing in AB 2655 prevents the enjoinder of speech through a temporary restraining order or preliminary injunction alternative to or in addition to suits under Sections 20515(b) and 20516; (ii) AB 2655 mandates the immediate removal of speech, without a determination that it is unprotected, so long as it is “substantially similar” to speech “previously removed” under the statute, see § 20513(c); and (iii) the statute acts as an

1 overarching prior restraint by, in its pursuit of eliminating
2 certain speech altogether, imposing a system of censorship that
3 requires platforms to remove the speech within 72 hours absent a
4 final ruling that it is unprotected.

5 101. Second, because AB 2655 imposes content-, viewpoint-,
6 and speaker-based speech restrictions, it triggers constitutional
7 review under strict scrutiny, which it cannot withstand.

8 102. Covered platforms “present[] a curated and ‘edited
9 compilation of [third party] speech,’” which “is itself protected
10 speech.” *Moody*, 144 S. Ct. at 2409 (quoting *Hurley*, 515 U.S. at
11 570); see also *id.* at 2401 (“A private party’s collection of third-
12 party content into a single speech product (the operators’
13 ‘repertoire’ of programming) is itself expressive, and intrusion
14 into that activity must be specially justified under the First
15 Amendment.”). Moreover, the First Amendment protects “both the
16 right to speak freely and the right to refrain from speaking at
17 all.” *Wooley v. Maynard*, 430 U.S. 705, 714 (1977).

18 103. By forcing covered platforms to remove and modify
19 particular speech that they may not otherwise remove or modify –
20 i.e., certain election-related “materially deceptive content” –
21 and to create a reporting requirement to facilitate such removal
22 and modification, AB 2655 forces covered platforms to “‘speak a
23 particular message’ that they would not otherwise speak, which
24 constitutes compelled speech that dilutes their message.” *Kohls*,

2024 WL 4374134, at *5 (citing *Nat'l Inst. of Fam. & Life Advocs. v. Becerra* ("NIFLA"), 585 U.S. 755, 766 (2018); *X Corp. v. Bonta*, 116 F.4th 888, 900–02 (9th Cir. 2024)); see also *Washington Post v. McManus*, 944 F.3d 506, 511–13, 519 (4th Cir. 2019) (striking down state law that required, in an effort to address foreign interference in U.S. elections, "online platforms," within "48 hours of an ad being purchased," to "display somewhere on their site the identity of the purchaser, the individuals exercising control over the purchaser, and the total amount paid for the ad," and declaring the law "a compendium of traditional First Amendment infirmities" that would "chill speech"); *id.* at 515 ("each banner feature of the Act – the fact that it is content-based, targets political expression, and compels certain speech – poses a real risk of either chilling speech or manipulating the marketplace of ideas"). AB 2655 also impermissibly substitutes the judgment of the government for that of covered platforms as to what constitutes "materially deceptive content" covered by the statute and whether it should remain on their platforms.

104. In addition, the underlying content that AB 2655 targets – i.e., the content delineated in §§ 20513(a) and 20514(a) – is itself constitutionally protected. In other words, AB 2655 is not merely a "restriction on knowing falsehoods that fall outside of the category of false speech protected by the First Amendment as

1 articulated in” *Alvarez*, 567 U.S. 709. *Kohls*, 2024 WL 4374134, at
2 *3.

3 105. Accordingly, AB 2655 is a content-based law – that is,
4 it “target[s] speech based on its communicative content,” *Reed v.*
5 *Town of Gilbert, Ariz.*, 576 U.S. 155, 163 (2015) – and no exception
6 applies here to the longstanding rule that such regulations trigger
7 strict scrutiny. *NIFLA*, 585 U.S. at 767 (quoting *Brown v.*
8 *Entertainment Merchants Assn.*, 564 U.S. 786, 792 (2011)) (“This
9 Court’s precedents do not permit governments to impose content-
10 based restrictions on speech without persuasive evidence . . . of
11 a long (if heretofore unrecognized) tradition to that effect.”).
12 By “specifically target[ing] speech within political or electoral
13 content pertaining to candidates, electoral officials, and other
14 election communication,” AB 2655 “delineates acceptable and
15 unacceptable content based on its purported truth or falsity and
16 is an archetypal content-based regulation that our constitution
17 considers dubious and subject to strict scrutiny.” *Kohls*, 2024 WL
18 4374134, at *4.

19 106. AB 2655 triggers strict scrutiny for two additional
20 reasons. First, AB 2655 discriminates based on the identity of
21 the speaker; it applies only to certain speakers (i.e., to covered
22 platforms such as X), while exempting others (e.g., certain
23 broadcasting stations, online newspapers, and magazines). See,
24 e.g., *Sorrell v. IMS Health Inc.*, 564 U.S. 552, 571 (2011) (laws

1 that interfere with the speech rights of only certain speakers
2 “justify application of heightened scrutiny” particularly when they
3 are aimed at specific content); see also *Moody*, 144 S. Ct. at 2405
4 (quoting *Tornillo*, 418 U.S. at 258) (“‘The choice of material,’
5 the ‘decisions made [as to] content,’ the ‘treatment of public
6 issues’ – ‘whether fair or unfair’ – all these ‘constitute the
7 exercise of editorial control and judgment.’ . . . **For a paper,**
8 **and for a platform too.**”). Second, AB 2655 discriminates based on
9 viewpoint, because it permits election-related content that is
10 “‘positive’ about a person,” while restricting such content if it
11 is “derogatory.” *Iancu v. Brunetti*, 588 U.S. 388, 393 (2019)
12 (quoting *Matal v. Tam*, 582 U.S. 218, 249 (2017) (Kennedy, J.,
13 concurring)) (explaining that such differential treatment
14 “reflects the Government’s disapproval of a subset of messages it
15 finds offensive” and “is the essence of viewpoint discrimination”).

16 107. AB 2655 may stand, then, only if the government proves
17 that the statute is “narrowly tailored to serve compelling state
18 interests,” *NIFLA*, 585 U.S. at 766 (quoting *Reed*, 576 U.S. at 163),
19 and no “less restrictive alternative would serve the [g]overnment’s
20 purpose,” *X Corp.*, 116 F.4th at 903 (quoting *United States v.*
21 *Playboy Entm’t Grp., Inc.*, 529 U.S. 803, 813 (2000)).

22 108. AB 2655 fails strict scrutiny because, even if California
23 has a compelling interest in protecting free and fair elections,
24 AB 2655 is not the “least restrictive means available for advancing

1 [that] interest," *Kohls*, 2024 WL 4374134, at *4 (quoting *NetChoice*,
2 *LLC*, 113 F.4th at 1121), and the "First Amendment does not 'permit
3 speech-restrictive measures when the state may remedy the problem
4 by implementing or enforcing laws that do not infringe on speech,'" *id.* (quoting *IMDb.com, Inc. v. Becerra*, 962 F.3d 1111, 1125 (9th
5 Cir. 2020)); see also *Ex parte Stafford*, 2024 WL 4031614, at *4-6
6 (Tex. Crim. App. Sept. 4, 2024) (applying strict scrutiny and
7 striking down on First Amendment grounds Texas statute prohibiting
8 "knowingly represent[ing] in a campaign communication that the
9 communication emanates from a source other than its true source"
10 because there were "narrower means of achieving the State
11 interests," including enforcing an existing statute). Moreover,
12 it is not a "valid, let alone substantial" interest for a state to
13 seek "to correct the mix of speech" that "social-media platforms
14 present." *Moody*, 144 S. Ct. at 2407; see also *id.* at 2409 (quoting
15 *Pac. Gas & Elec. Co.*, 475 U.S. at 20) ("[A] State 'cannot advance
16 some points of view by burdening the expression of others.'").²²

18 109. AB 2655 is facially invalid under the First Amendment
19 because "a substantial number of [the law's] applications are
20 unconstitutional, judged in relation to the statute's plainly
21 legitimate sweep." *Americans for Prosperity Foundation v. Bonta*,
22 594 U.S. 595, 615 (2021). It is also unconstitutional as-applied
23 to X Corp. specifically.

24

²² Nor would AB 2655 survive under any lesser standard of review.

1 110. There is a *bona fide* and actual controversy between X
2 Corp. and Defendants because Defendants are charged with enforcing,
3 and intend to enforce, AB 2655, even though it violates the First
4 Amendment of the United States Constitution and Article I, Section
5 2, of the California Constitution, both facially and as-applied to
6 X Corp.

7 111. X Corp. maintains that AB 2655 is illegal and
8 unconstitutional. Defendants claim otherwise.

9 112. X Corp. requests a judicial determination regarding the
10 validity of AB 2655 to prevent the harm caused by its enactment.
11 Such a determination is both necessary and appropriate to avoid
12 the deprivation of X's and the other covered platforms'
13 constitutional rights, which would occur if AB 2655 is applied to
14 X Corp. or any other covered platform.

15 113. Given the violation of the First Amendment of the United
16 States Constitution and Article I, Section 2, of the California
17 Constitution, X Corp. seeks preliminary and permanent injunctive
18 relief against enforcement of AB 2655. X and the other covered
19 platforms would be irreparably harmed if they were forced to comply
20 with AB 2655's requirements and have no adequate remedy at law.

SECOND CAUSE OF ACTION

**(Declaratory Relief and Preliminary and Permanent Injunctive
Relief for Immunity Under and Preemption by 47 U.S.C.
§§ 230(c)(1) and 230(c)(2))**

114. X Corp. realleges and incorporates herein by reference each and every allegation set forth above.

115. 47 U.S.C. §§ 230(c)(1) and 230(c)(2) each directly conflict with, and thus preempt, AB 2655.

116. 47 U.S.C. § 230(e)(3) provides that “[n]o cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section.”

117. AB 2655 imposes liability on covered platforms by holding them responsible for the content of what is on their platforms, as if they were the publisher of that content. It requires removal and labeling of content that the State disfavors (i.e., “materially deceptive content” that is otherwise covered by the statute) and requires removal and labeling of such content if the covered platforms fail to comply. See §§ 20513–20516.

118. “Liability” under Section 230(e)(3) includes being subjected to the kind of injunctive and other equitable relief authorized by AB 2655’s Enforcement Provisions. See, e.g., *Hassell v. Bird*, 5 Cal. 5th 522, 544–45 (2018) (finding that Section 230 barred “cause[s] of action” directing Yelp to remove defamatory consumer reviews).

119. X is an “interactive computer service,” as that term is defined under 47 U.S.C. § 230(f)(2).

1 **Section 230(c)(1)**

2 120. AB 2655 directly contravenes the immunity provided to
3 the covered platforms by 47 U.S.C. § 230(c)(1), which prohibits
4 treating interactive computer service providers as the “publisher
5 or speaker of any information provided by another information
6 content provider.”

7 121. AB 2655’s Enforcement Provisions violate Section
8 230(c)(1) because they provide causes of action for “injunctive or
9 other equitable relief against” the covered platform to remove or
10 (by adding a disclaimer) alter certain content posted on the
11 platform by its users. See §§ 20515(b), 20516. AB 2655 thus
12 treats covered platforms “as the publisher or speaker of any
13 information provided by another information content provider.” 47
14 U.S.C. § 230(c)(1).

15 122. Section 230(c)(1) bars such liability where the alleged
16 duty violated derives from an entity’s conduct as a “publisher,”
17 including “reviewing, editing, and deciding whether to publish or
18 withdraw from publication third-party content.” See, e.g., *Barnes*
19 *v. Yahoo!, Inc.*, 570 F.3d 1096, 1102 (9th Cir. 2009) (finding that
20 Yahoo! was entitled to immunity under Section 230(c)(1) from claims
21 concerning failure to remove offending profile), *as amended* (Sept.
22 28, 2009); *Calise v. Meta Platforms, Inc.*, 103 F.4th 732, 744 (9th
23 Cir. 2024) (finding that Meta was immune under Section 230(c)(1)

1 from claims that would require Meta to “actively vet and evaluate
2 third-party ads” in order to remove them).

3 **Section 230(c)(2)(B)**

4 123. AB 2655 also directly contravene the immunity provided
5 to the covered platforms by 47 U.S.C. § 230(c)(2)(B), which
6 prohibits holding interactive computer service providers “liable
7 on account of . . . any action taken to enable or make available
8 to information content providers or others the technical means to
9 restrict access to [objectionable] material.”

10 124. Section 20516 of AB 2655’s Enforcement Provisions
11 violates Section 230(c)(2)(B) because it provides causes of action
12 for “injunctive or other equitable relief against” covered
13 platforms that attempt to comply with the Reporting Requirement,
14 but do so in a manner that, in the government attorney’s view, does
15 not meet the reporting “require[ments]” of “subdivision (a) of
16 Section 20515.” § 20516.

17 125. A covered platform’s attempt to comply with the Reporting
18 Requirement (i.e., creating a reporting mechanism for users to
19 report content covered by the statute) is an action to make
20 available the technical means to restrict access to objectionable
21 content, as contemplated by Section 230(c)(2)(B), and covered
22 platforms will face enforcement if they do not comply to the
23 satisfaction of the California government.
24

1 126. There is a *bona fide* and actual controversy between X
2 Corp. and Defendants because Defendants are charged with enforcing,
3 and intend to enforce, AB 2655, even though such enforcement is
4 precluded and preempted by 47 U.S.C. §§ 230(c)(1) and 230(c)(2).

5 127. X Corp. maintains that AB 2655 is invalid and void as a
6 matter of law. Defendants claim otherwise.

7 128. X Corp. seeks a declaratory judgment that AB 2655 is
8 legally invalid and unenforceable because it is precluded and
9 preempted by 47 U.S.C. §§ 230(c)(1) and 230(c)(2).

10 129. Given the violation of 47 U.S.C. §§ 230(c)(1) and
11 230(c)(2), X Corp. seeks preliminary and permanent injunctive
12 relief against enforcement of AB 2655. X Corp. would be irreparably
13 harmed if it were forced to comply with, or litigate, AB 2655's
14 requirements and has no adequate remedy at law.

15 **THIRD CAUSE OF ACTION**

16 **(Declaratory Relief and Preliminary and Permanent Injunctive**
17 **Relief for Violations of the First and Fourteenth Amendments of**
the United States Constitution (42 U.S.C. § 1983) for Vagueness)

18 130. X Corp. realleges and incorporates herein by reference
19 each and every allegation set forth above.

20 131. AB 2655 is void for vagueness under the First and
21 Fourteenth Amendments of the U.S. Constitution because the
22 statute's requirements and prohibitions are so unintelligible that
23 X and the other covered platforms cannot understand what the law
24 prohibits.

1 132. X and the other covered platforms cannot understand what
2 would constitute a “[d]eepfake” under Section 20512(d) because they
3 cannot understand what “would falsely appear to a reasonable person
4 to be an authentic record of the actual speech or conduct of the
5 individual depicted in the media.”

6 133. X and the other covered platforms cannot understand what
7 would constitute “[m]aterially deceptive content” under Section
8 20512(i) because they cannot understand what “would falsely appear
9 to a reasonable person to be an authentic record of the content
10 depicted in the media.”

11 134. X and the other covered platforms cannot understand what
12 would constitute “state-of-the-art techniques” under Sections
13 20513(a), 20513(c), and 20514(a).

14 135. X and the other covered platforms cannot understand what
15 would be “reasonably likely to harm the reputation or electoral
16 prospects of a candidate” under Section 20513(a)(2)(A).

17 136. X and the other covered platforms cannot understand what
18 would be “reasonably likely to falsely undermine confidence in the
19 outcome of one or more election contests” under Sections
20 20513(a)(2)(B) and 20513(a)(2)(C).

21 137. X and the other covered platforms cannot understand what
22 would “influence[] an election in California” under Section
23 20513(a)(2)(C).

24

1 138. X and the other covered platforms cannot understand what
2 would constitute a candidate for elective office, an elections
3 official, or an elected official being “portrayed as doing or
4 saying something” that they “did not do or say” under Sections
5 20513(a)(2)(A), 20513(a)(2)(B), and 20513(a)(2)(C).

6 139. X and the other covered platforms cannot understand what
7 would constitute an “easy-to-understand format” under Section
8 20514(d).

9 140. Due to the vagueness and ambiguity of these terms and
10 phrases, AB 2655 fails to give X and the other covered platforms
11 “a reasonable opportunity to know what” the statute “prohibit[s].”
12 *Hunt v. City of Los Angeles*, 638 F.3d 703, 712 (9th Cir. 2011).

13 141. AB 2655 “impermissibly delegates basic policy matters to
14 policemen, judges, and juries for resolution on an *ad hoc* and
15 subjective basis, with the attendant dangers of arbitrary and
16 discriminatory application.” *Id.*; see also, e.g., *NAACP v. Button*,
17 371 U.S. 415, 432 (1963) (holding that the “standards of
18 permissible statutory vagueness are strict in the area of free
19 expression”).

20 142. There is a *bona fide* and actual controversy between X
21 Corp. and Defendants because Defendants are charged with enforcing,
22 and intend to enforce, AB 2655, even though it violates the First
23 and Fourteenth Amendments of the United States Constitution for
24 vagueness.

143. X Corp. maintains that AB 2655 is illegal and unconstitutional. Defendants claim otherwise.

144. X Corp. requests a judicial determination regarding the validity of AB 2655 to prevent the harm caused by its enactment. Such a determination is both necessary and appropriate to avoid the deprivation of X's and the other covered platforms' constitutional rights, which would occur if AB 2655 is applied to X or any other covered platform.

145. Given the violation of the First and Fourteenth Amendments of the United States for vagueness, X Corp. seeks preliminary and permanent injunctive relief against enforcement of AB 2655. X and the other covered platforms would be irreparably harmed if they were forced to comply with AB 2655's requirements and have no adequate remedy at law.

PRAYER FOR RELIEF

WHEREFORE, X Corp. respectfully requests that this Court enter judgment in X Corp.'s favor and grant the following relief:

1. A declaration that AB 2655 violates the First Amendment of the United States Constitution and Article I, Section 2, of the California Constitution, both facially and as-applied to X Corp.;

2. A declaration that the injunctive and other equitable relief provided by AB 2655 is precluded and preempted by 47 U.S.C. §§ 230(c)(1) and 230(c)(2) and is therefore null and void and has no legal effect;

1 3. A declaration that AB 2655 violates the First and
2 Fourteenth Amendments of the United States Constitution for
3 vagueness;

4 4. A preliminary and permanent injunction enjoining
5 Defendants and their employees, agents, and successors in office
6 from enforcing AB 2655;

7 5. An award of fees, costs, expenses, and disbursements,
8 including attorneys' fees, to which X Corp. is entitled pursuant
9 to 42 U.S.C. § 1988 and other applicable law; and

10 6. Such other and further relief as the Court deems just
11 and proper.

12 **DEMAND FOR JURY TRIAL**

13 Pursuant to Federal Rule of Civil Procedure 38, X Corp.
14 demands a trial by jury in this action of all issues so triable.
15
16
17
18
19
20
21
22
23
24

1 Dated: November 14, 2024

2
3 By: /s/ William R. Warne
DOWNEY BRAND LLP
4 William R. Warne (SBN 141280)
Meghan M. Baker (SBN 243765)
5 621 Capitol Mall, 18th Floor
Sacramento, CA 95814
6 Phone: 916-444-1000
Facsimile: 916-520-5910

7 CAHILL GORDON & REINDEL LLP
Joel Kurtzberg (*pro hac vice pending*, SBN NY 1758184)
8 Floyd Abrams (*pro hac vice pending*, SBN NY 2835007)
Jason Rozbruch (*pro hac vice pending*, SBN NY 5753637)
32 Old Slip
9 New York, NY 10005
Phone: 212-701-3120
10 Facsimile: 212-269-5420
jkurtzberg@cahill.com
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Exhibit 1



Post



Eliot Higgins
@EliotHiggins



Making pictures of Trump getting arrested while waiting for Trump's arrest.



5:22 PM · Mar 20, 2023 · **6.9M** Views

4,973 Reposts **2,317** Quotes **37.5K** Likes **3,092** Bookmarks



3K



New to X?

Sign up now to get your own personalized timeline!



Sign up with Google



Sign up with Apple

Create account

By signing up, you agree to the [Terms of Service](#) and [Privacy Policy](#), including [Cookie Use](#).

Something went wrong. Try reloading.



Retry

[Terms of Service](#) [Privacy Policy](#) [Cookie Policy](#)
[Accessibility](#) [Ads info](#) [More ...](#) © 2024 X Corp.

Don't miss what's happening

People on X are the first to know.

Log in

Sign up

Exhibit 2

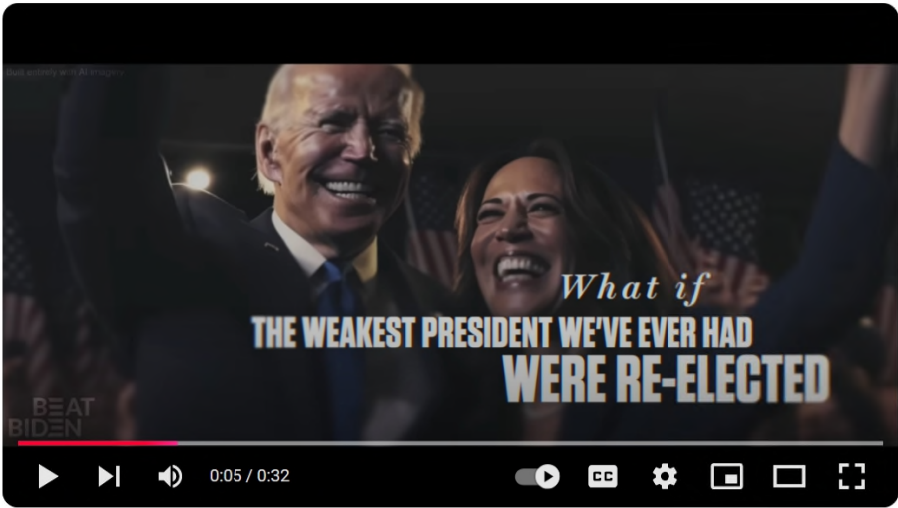


YouTube

Search



Sign in



Beat Biden



GOP

92.7K subscribers

Subscribe

1.9K



Share



347,247 views Apr 25, 2023

An AI-generated look into the country's possible future if Joe Biden is re-elected in 2024.

Click below to subscribe now!

http://www.youtube.com/c/GOP?sub_conf...

<http://www.GOP.com>

How this content was made

Altered or synthetic content

Sound or visuals were significantly edited or digitally generated. [Learn more](#)

Transcript

Follow along using the transcript.

Show transcript



GOP

92.7K subscribers

Videos

About

Show less

1,276 Comments

Sort by



Add a comment...



Gov. Andy Beshear Victory Speech | Primary Election 202...

KET - Kentucky Educational Television
22K views • 1 year ago



Irish CNN reporter sees how hometown is helping Ukraine...

CNN
53K views • 2 years ago



First Debate Cold Open - SNL

Saturday Night Live
37M views • 4 years ago



Steve Kornacki on when to expect to see results on...

MSNBC
644K views • 1 day ago
New



How the Republican Party went from Lincoln to Trump

Vox
10M views • 8 years ago



Ronald Reagan's one-liners

CBS Sunday Morning
7.2M views • 10 years ago



JUST IN: RFK Jr. Details Work He Wants To Do If Trump Wins...

Forbes Breaking News
271K views • 4 days ago
New



Thank You, America.

Joe Biden
478K views • 2 months ago



'When I Hear Kamala Speak...': Hulk Hogan Roasts 'Bad...

Forbes Breaking News
941K views • 8 days ago



President Joe Biden's most embarrassing moments | 2023...

7NEWS Australia
1.1M views • 11 months ago



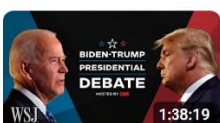
Fox News Kamala Harris Interview Cold Open - SNL

Saturday Night Live
8M views • 2 weeks ago



Trump Stops Michigan Rally Speech To Play Video To...

Forbes Breaking News
165K views • 3 days ago
New



Full Debate: Biden and Trump in the First 2024 Presidential...

The Wall Street Journal
22M views • Streamed 4 months ago

Exhibit 3

SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2023-2024 Regular Session

AB 2655 (Berman)
Version: June 11, 2024
Hearing Date: July 2, 2024
Fiscal: Yes
Urgency: No
CK

SUBJECT

Defending Democracy from Deepfake Deception Act of 2024

DIGEST

This bill establishes the Defending Democracy from Deepfake Deception Act of 2024, which requires a large online platform to block the posting or sending of materially deceptive and digitally modified or created content related to elections, during specified periods before and after an election. It requires these platforms to label certain additional content inauthentic, fake, or false during specified periods before and after an election and to provide mechanisms to report such content.

EXECUTIVE SUMMARY

The rapid advancement of AI technology, specifically the wide-scale introduction of generative AI models, has made it drastically cheaper and easier to produce synthetic content – audio, images, text, and video recordings that are not real, but that are so realistic that they are virtually impossible to distinguish from authentic content, including so-called “deepfakes.” In the context of election campaigns, such deepfakes can be weaponized to deceive voters into thinking that a candidate said or did something which the candidate did not, or otherwise falsely call election results into question. A series of bills currently pending before this Committee attempt to address these issues by restricting or labeling AI-altered or –generated content. However, this bill specifically targets social media platforms and such materially deceptive content on their platforms, requiring platforms to block and prevent it, label it, and provide mechanisms for reporting it.

The bill is sponsored by the California Initiative for Technology & Democracy. It is supported by various organizations, including the League of Women Voters of California and Disability Rights California. It is opposed by Oakland Privacy and various industry associations, including TechNet. The bill passed out of the Senate Elections and Constitutional Amendments Committee on a 6 to 1 vote.

PROPOSED CHANGES TO THE LAW

Existing law:

- 1) Provides that “Congress shall make no law... abridging the freedom of speech...” (U.S. Const., amend. 1.)
- 2) Applies the First Amendment to the states through operation of the Fourteenth Amendment. (*Gitlow v. New York* (1925) 268 U.S. 652; *NAACP v. Alabama* (1925) 357 U.S. 449.)
- 3) Provides, in federal law, that a provider or user of an interactive computer service shall not be treated as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. § 230(c)(1).)
- 4) Provides that a provider or user of an interactive computer service shall not be held liable on account of:
 - a. any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
 - b. any action taken to enable or make available to information content providers or others the technical means to restrict access to such material. (47 U.S.C. § 230(c)(2).)
- 5) Provides that no provider or user of an interactive computer service shall be treated for liability purposes as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. § 230.)
- 6) Defines “materially deceptive audio or visual media” as an image or an audio or video recording of a candidate’s appearance, speech, or conduct that has been intentionally manipulated in a manner such that both of the following conditions are met:
 - a. The image or audio or video recording would falsely appear to a reasonable person to be authentic.
 - b. The image or audio or video recording would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than that person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording. (Elec. Code § 20010(e).)

- 7) Prohibits a person, committee, or other entity from distributing with actual malice materially deceptive audio or visual media of a candidate with the intent to injure the candidate's reputation or to deceive a voter into voting for or against the candidate within 60 days of an election at which a candidate for elective office will appear on the ballot, as specified and unless specified conditions are met. (Elec. Code § 20010(a).)
- 8) Exempts audio or visual media that includes a disclosure stating: "This _____ has been manipulated." Requires the blank in the disclosure to be filled with a term that most accurately describes the media, as specified. Requires the following disclosures for visual and audio-only media:
 - a. For visual media, the text of the disclosure shall appear in a size that is easily readable by the average viewer and no smaller than the largest font size of other text appearing in the visual media. If the visual media does not include any other text, then the disclosure shall appear in a size that is easily readable by the average viewer. Requires, for visual media that is video, the disclosure to be displayed throughout the duration of the video.
 - b. For audio-only media, the disclosure shall be read in the clearly spoken manner and in a pitch that can be easily heard by the average listener, at the beginning of the audio, at the end of the audio, and, if the audio is greater than two minutes in length, interspersed within the audio at intervals of not greater than two minutes each. (Elec. Code § 20010(b).)
- 9) Permits a candidate for elective office whose voice or likeness appears in a materially deceptive audio or visual media distributed in violation of the above provisions, to seek injunctive or other equitable relief prohibiting the distribution of audio or visual media in violation. (Elec. Code § 20010(c)(1).)
- 10) Permits a candidate for elective office whose voice or likeness appears in materially deceptive audio or visual media distributed in violation of the above provisions to bring an action for general or special damages against the person, committee, or other entity that distributed the materially deceptive audio or visual media, as specified. Requires the plaintiff to bear the burden of establishing the violation through clear and convincing evidence in any civil action alleging a violation, as specified. (Elec. Code § 21101(c)(2).)

This bill:

- 1) Establishes the Defending Democracy from Deepfake Deception Act of 2024.
- 2) Requires a large online platform, using state-of-the-art, best available tools to detect materially deceptive content, to develop and implement procedures for blocking and preventing, and, if the platform knows or should know that the materially deceptive content meets the requirements hereof, to block and prevent

the posting or sending of any materially deceptive content, if all of the following conditions are met:

- a) The content is posted or sent during a period beginning 120 days before the election and through the day of the election. For content that depicts or pertains to elections officials, this period shall extend to the 60th day after the election.
 - b) The materially deceptive content is any of the following:
 - i. A candidate for elective office portrayed as doing or saying something that the candidate did not do or say and that is reasonably likely to harm the reputation or electoral prospects of a candidate.
 - ii. An elections official portrayed as doing or saying something in connection with the performance of their elections-related duties that the elections official did not do or say and that is reasonably likely to falsely undermine confidence in the outcome of one or more election contests.
 - iii. An elected official portrayed as doing or saying something that influences the election that the elected official did not do or say and that is reasonably likely to falsely undermine confidence in the outcome of one or more election contests.
 - c) The person or entity who created the materially deceptive content did so knowing it was false or with reckless disregard for the truth. There shall be a rebuttable presumption that the person or entity knew the materially deceptive content was false or acted with reckless disregard for the truth if the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.
- 3) Requires, notwithstanding the above, a large online platform to allow a candidate for elective office, during a period beginning 120 days before the election and through the day of the election, to portray themselves as doing or saying something that the candidate did not do or say, but only if the digital content includes a disclosure meeting specified conditions and states the following: "This [category of content] has been manipulated."
- 4) Requires a large online platform, using state-of-the-art, best available tools to detect materially deceptive content to develop and implement procedures for labeling such content as inauthentic, fake, or false if all of the following conditions are met:
- a) The materially deceptive content is either of the following:
 - i. Meets the standards set above, but is posted or sent outside the applicable time period.
 - ii. Appears within an advertisement or election communication and is not subject to the above.

- b) The person or entity who created the materially deceptive content did so knowing it was false or with reckless disregard for the truth. There shall be a rebuttable presumption that the person or entity knew the materially deceptive content was false or acted with reckless disregard for the truth if the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.
 - c) The large online platform knows or should know that the materially deceptive content meets the requirements of this section.
- 5) Specifies required functionality of the label above and states the labeling requirement applies during redistricting and during a period from one year before the election through election day. If the content involves elections officials, the electoral college process, the canvass of the vote, or election-related equipment or property, the time period is extended 60 days beyond the election.
- 6) Requires a large online platform to provide an easily accessible way for California residents to report to that platform content subject to the above provisions that was not blocked or labeled as required. The online platform shall respond to the person who made the report, within 36 hours of the report, describing any action taken or not taken by the online platform with respect to the content.
- 7) Authorizes a candidate for elective office, elected official, or elections official who has made a report and who either has not received a response within 36 hours or disagrees with the response, as well as the Attorney General or any district attorney or city attorney, to seek injunctive or other equitable relief against the online platform to compel compliance. The plaintiff shall bear the burden of establishing the violation through clear and convincing evidence. The court is required to award a prevailing plaintiff reasonable attorney's fees and costs. Such actions are given precedence in accordance with Section 35 of the Code of Civil Procedure.
- 8) Clarifies that it applies to materially deceptive content, regardless of the language used in the content. If the language used is not English, the required disclosure and label must appear in the language used as well as in English.
- 9) Requires a large online platform that blocks or labels any materially deceptive content to maintain a copy of the digital content for a period of not less than five years from the election and shall make such digital content available to the Secretary of State, the Fair Political Practices Commission, and researchers, if requested.

- 10) Exempts from the scope of the bill the following:
 - a) A regularly published online newspaper, magazine, or other periodical of general circulation that routinely carries news and commentary of general interest, and that publishes any materially deceptive content that an online platform is required to block or label based on this chapter, if the publication contains a clear disclosure that the materially deceptive content does not accurately represent any actual event, occurrence, appearance, speech, or expressive conduct.
 - b) Materially deceptive content that constitutes satire or parody.
- 11) Includes findings and declarations and a severability clause.
- 12) Defines the relevant terms, including:
 - a) “Materially deceptive content” means audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media.
 - b) “Large online platform” means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.

COMMENTS

1. Blurring reality: AI-generated content

Generative AI is a type of artificial intelligence that can create new content, including text, images, code, or music, by learning from existing data. Generative AI models can produce realistic and novel artifacts that resemble the data they were trained on, but do not copy it. For example, generative AI can write a poem, draw a picture, or compose a song based on a given prompt or theme. Generative AI enables users to quickly generate new content based on a variety of inputs. Generative AI models use neural networks to identify the patterns and structures within existing data to generate new and original content.

The world has been in awe of the powers of this generative AI since the widespread introduction of AI systems such as ChatGPT. However, the capabilities of these advanced systems leads to a blurring between reality and fiction. The Brookings Institution lays out the issue:

Over the last year, generative AI tools have made the jump from research prototype to commercial product. Generative AI models like OpenAI’s ChatGPT and Google’s Gemini can now generate realistic text and images that are often indistinguishable from human-authored content, with

generative AI for audio and video not far behind. Given these advances, it's no longer surprising to see AI-generated images of public figures go viral or AI-generated reviews and comments on digital platforms. As such, generative AI models are raising concerns about the credibility of digital content and the ease of producing harmful content going forward.

Against the backdrop of such technological advances, civil society and policymakers have taken increasing interest in ways to distinguish AI-generated content from human-authored content.¹

One expert at the Copenhagen Institute for Future Studies estimates that should large generative-AI models run amok, up to 99 percent of the internet's content could be AI-generated by 2025 to 2030.² The problematic applications are seemingly infinite, whether it be deepfakes to blackmail or shame victims, false impersonations to commit fraud, or other nefarious purposes. Infamously, in January of this year, Taylor Swift was the victim of sexually explicit, nonconsensual deepfake images using AI that were widely spread across social media platforms.³ Perhaps more disturbingly, a trend has emerged in schools of students creating such images: "At schools across the country, people have used deepfake technology combined with real images of female students to create fraudulent images of nude bodies. The deepfake images can be produced using a cellphone."⁴ As more of the population becomes aware of the potential to realistically fake images, video, and text, some will use the skepticism that creates to challenge the authenticity of real content, a phenomena coined the "liar's dividend."⁵

Relevant here, AI and specifically generative AI can spread misinformation regarding elections with ease, both in California and across the world:

Artificial intelligence is supercharging the threat of election disinformation worldwide, making it easy for anyone with a smartphone

¹ Siddarth Srinivasan, *Detecting AI fingerprints: A guide to watermarking and beyond* (January 4, 2024) Brookings Institution, <https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/#:~:text=Google%20also%20recently%20announced%20SynthID,model%20to%20detect%20the%20watermark>. All internet citations are current as of June 23, 2024.

² Lonnie Lee Hood, *Experts Say That Soon, Almost The Entire Internet Could Be Generated by AI* (March 4, 2022) The Byte, <https://futurism.com/the-byte/ai-internet-generation>.

³ Brian Contreras, *Tougher AI Policies Could Protect Taylor Swift – And Everyone Else – From Deepfakes* (February 8, 2024) Scientific American, <https://www.scientificamerican.com/article/tougher-ai-policies-could-protect-taylor-swift-and-everyone-else-from-deepfakes/>.

⁴ Hannah Fry, Laguna Beach High School investigates 'inappropriate' AI-generated images of students (April 2, 2024) Los Angeles Times, <https://www.latimes.com/california/story/2024-04-02/laguna-beach-high-school-investigating-creation-of-ai-generated-images-of-students>.

⁵ Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* (July 14, 2018) 107 California Law Review 1753 (2019), <https://ssrn.com/abstract=3213954>.

and a devious imagination to create fake – but convincing – content aimed at fooling voters.

It marks a quantum leap from a few years ago, when creating phony photos, videos or audio clips required teams of people with time, technical skill and money. Now, using free and low-cost generative artificial intelligence services from companies like Google and OpenAI, anyone can create high-quality “deepfakes” with just a simple text prompt.

A wave of AI deepfakes tied to elections in Europe and Asia has coursed through social media for months, serving as a warning for more than 50 countries heading to the polls this year.

“You don’t need to look far to see some people ... being clearly confused as to whether something is real or not,” said Henry Ajder, a leading expert in generative AI based in Cambridge, England.

The question is no longer whether AI deepfakes could affect elections, but how influential they will be, said Ajder, who runs a consulting firm called Latent Space Advisory.

As the U.S. presidential race heats up, FBI Director Christopher Wray recently warned about the growing threat, saying generative AI makes it easy for “foreign adversaries to engage in malign influence.”⁶

On that last note, in February of this year, voters in New Hampshire received robocalls that are purported to have used an AI voice resembling President Joe Biden advising them against voting in the presidential primary and saving their vote for the November general election.⁷ The examples are endless:

Former President Donald Trump, who is running in 2024, has shared AI-generated content with his followers on social media. A manipulated video of CNN host Anderson Cooper that Trump shared on his Truth Social platform on Friday, which distorted Cooper’s reaction to the CNN town hall this past week with Trump, was created using an AI voice-cloning tool.

⁶ Ali Swenson & Kelvin Chan, *Election disinformation takes a big leap with AI being used to deceive worldwide* (March 14, 2024) Associated Press, <https://apnews.com/article/artificial-intelligence-elections-disinformation-chatgpt-bc283e7426402f0b4baa7df280a4c3fd>.

⁷ Em Steck & Andrew Kaczynski, *Fake Joe Biden robocall urges New Hampshire voters not to vote in Tuesday’s Democratic primary* (January 22, 2024) CNN, <https://www.cnn.com/2024/01/22/politics/fake-joe-biden-robocall/index.html>.

A dystopian campaign ad released last month by the Republican National Committee offers another glimpse of this digitally manipulated future. The online ad, which came after President Joe Biden announced his reelection campaign, and starts with a strange, slightly warped image of Biden and the text “What if the weakest president we’ve ever had was re-elected?”

A series of AI-generated images follows: Taiwan under attack; boarded up storefronts in the United States as the economy crumbles; soldiers and armored military vehicles patrolling local streets as tattooed criminals and waves of immigrants create panic.

“An AI-generated look into the country’s possible future if Joe Biden is re-elected in 2024,” reads the ad’s description from the RNC.

The RNC acknowledged its use of AI, but others, including nefarious political campaigns and foreign adversaries, will not, said Petko Stoyanov, global chief technology officer at Forcepoint, a cybersecurity company based in Austin, Texas. Stoyanov predicted that groups looking to meddle with U.S. democracy will employ AI and synthetic media as a way to erode trust.⁸

Legislatures across the country are pushing legislation that would address this looming threat.

2. Materially deceptive content in political advertisements

According to the author:

AB 2655 will ensure that online platforms restrict the spread of election-related deceptive deepfakes meant to prevent voters from voting or to deceive them based on fraudulent content. Deepfakes are a powerful and dangerous tool in the arsenal of those that want to wage disinformation campaigns, and they have the potential to wreak havoc on our democracy by attributing speech and conduct to a person that is false or that never happened. Advances in technology make it easy for practically anyone to generate this deceptive content, making it that much more important that we identify and restrict its spread before it has the chance to deceive voters and undermine our democracy.

⁸ David Klepper & Ali Swenson, *AI-generated disinformation poses threat of misleading voters in 2024 election* (May 14, 2023) PBS News, <https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election>.

Unlike existing law or other bills pending before this Committee in this area, this bill seeks to place responsibility on large online platforms with regard to “materially deceptive content” regarding elections, placing a series of obligations on them. The bill defines “materially deceptive content” as audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media. “Large online platform” means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.⁹

a. Preventing and blocking materially deceptive content

The bill requires platforms to develop and implement procedures for blocking and preventing, and, to block and prevent the posting or sending of, materially deceptive content, if the platform knows or should know that the materially deceptive content meets the requirements of the bill and certain conditions are met.

The materially deceptive content must portray one of the following. First is content portraying a candidate for elective office as doing or saying something they did not do or say and that is reasonably likely to harm the reputation or electoral prospects of a candidate. Or it must portray an elected official as doing or saying something that influences the election or an elections official as doing or saying something in connection with the performance of their elections-related duties that the official did not do or say and that is reasonably likely to falsely undermine confidence in the outcome of one or more election contests.

Second, the content must be posted or sent during a period beginning 120 days before the election and through the day of the election. For content that depicts or pertains to elections officials, this period shall extend to the 60th day after the election.

Finally, to trigger the requirement for platforms to block and prevent the content, the person or entity who created the content must have done so knowing it was false or with reckless disregard for the truth.

In any ensuing litigation, the bill establishes a rebuttable presumption that the person or entity knew the materially deceptive content was false or acted with reckless disregard for the truth if the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.

⁹ The author has agreed to an amendment that cross-references the existing definition for “social media platform, to replace the reference in the bill to “social network.”

Carved out of this obligation is digital content that a candidate for elective office posts or shares that portrays themselves as doing or saying something that they did not do, so long as there is a disclosure indicating that the content has been manipulated that meets certain specifications. However, not only is this content not subject to the requirement for a platform to block and prevent, but platforms are required to host such content and cannot prevent such material, with no exception for whether it violates the platform's terms and services. This provision similarly applies during the period starting 120 days before an election through election day.

b. Labeling materially deceptive content

Large online platforms are also required to develop and implement procedures for labeling materially deceptive content as inauthentic, fake, or false. This applies to such content that meets the requirements from the above section but falls outside of the specified time range or that does not meet the requirements but that appears within an advertisement or election communication. The person or entity who created the materially deceptive content must have also done so knowing it was false or with reckless disregard for the truth. The rebuttable presumption again applies. And, for this obligation to trigger, the large online platform must have known or should have known that the materially deceptive content meets these requirements.

The label must allow users to click or tap on it and to inspect all available provenance data about the content in an easy-to-understand format. The labeling requirement applies during specified time periods: (1) the period starting one year before the election through election day; (2) that period through the 60th day after the election, if it depicts or pertains to elections officials, the electoral college process, a voting machine, ballot, voting site, or other property or equipment related to an election, or the canvass of the vote; and (3) during a governmental process related to redistricting, as provided.

c. Retention requirement

Content that either requires such a label or that must be blocked and prevented from being posted and shared must be retained by the platform for not less than five years from the relevant election. Platforms must share the content, upon request, with the Secretary of State, the Fair Political Practices Commission, and researchers.

d. Reporting mechanism

Lastly, the bill requires a large online platform to provide an easily accessible way for California residents to report to that platform content subject to the above provisions that was not blocked or labeled as required. The online platform shall respond to the person who made the report, within 36 hours, describing any action taken or not taken by the online platform.

e. Enforcement

The bill provides standing to candidates, elected officials, or elections officials who have made reports but who have either not received a timely response or who disagree with it to bring an action for injunctive and other equitable relief. The Attorney General, district attorneys, and city attorneys are also so authorized. A prevailing plaintiff is entitled to attorneys' fees and costs. Such actions are given precedence in the courts.

However, plaintiffs in such actions are required to establish a violation by clear and convincing evidence.

3. Legal concerns

Concerns have been raised about whether the bill runs afoul of federal statutory and constitutional law. Namely, whether the bill is preempted by Section 230 of the Communications Decency Act, 47 U.S.C. § 230 and the First Amendment to the United States Constitution.

a. Section 230

Section 230 does not apply to the *users* of social media (or the internet generally), but rather applies to the *platforms themselves*. In the early 1990s, prior to the enactment of Section 230, two trial court orders – one in the United States District Court for the Southern District of New York, and New York state court – suggested that internet platforms could be held liable for allegedly defamatory statements made by the platforms' users if the platforms engaged in any sort of content moderation (e.g., filtering out offensive material).¹⁰ In response, two federal legislators and members of the burgeoning internet industry crafted a law that would give internet platforms immunity from liability for users' statements, even if they might have reason to know that statements might be false, defamatory, or otherwise actionable.¹¹ The result – Section 230 – was relatively uncontroversial at the time, in part because of the relative novelty of the internet and in part because Section 230 was incorporated into a much more controversial internet regulation scheme that was the subject of greater debate.¹²

¹⁰ See *Cubby, Inc. v. Compuserve, Inc.* (S.D.N.Y. 1991) 776 F.Supp. 135, 141; *Stratton Oakmont v. Prodigy Servs. Co.* (N.Y. Sup. Ct., May 26, 1995) 1995 N.Y. Misc. LEXIS 229, *10-14. These opinions relied on case law developed in the context of other media, such as whether bookstores and libraries could be held liable for distributing defamatory material when they had no reason to know the material was defamatory. (See *Cubby, Inc.*, 776 F. Supp. at p. 139; *Smith v. California* (1959) 361 U.S. 147, 152-153.)

¹¹ Kosseff, *The Twenty-Six Words That Created The Internet* (2019) pp. 57-65.

¹² *Id.* at pp. 68-73. Section 230 was added to the Communications Decency Act of 1996 (title 5 of the Telecommunications Act of 1996, Pub. L. 104-104, 110 Stat. 56), which would have imposed criminal liability on internet platforms if they did not take steps to prevent minors from obtaining "obscene or indecent" material online. The Supreme Court invalidated the CDA, except for Section 230, on the basis that it violated the First Amendment. (See *Reno v. ACLU* (1997) 521 U.S. 844, 874.)

The crux of Section 230 is laid out in two parts. The first provides that “[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”¹³ The second provides a safe harbor for content moderation, by stating that no provider or user shall be held liable because of good-faith efforts to restrict access to material that is “obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.”¹⁴

Together, these two provisions give platforms immunity from any civil or criminal liability that could be incurred by user statements, while explicitly authorizing platforms to engage in their own content moderation without risking that immunity. Section 230 specifies that “[n]o cause of action may be brought and no liability may be imposed under any State law that is inconsistent with this section.”¹⁵ Courts have applied Section 230 in a vast range of cases to immunize internet platforms from “virtually all suits arising from third-party content.”¹⁶

This bill provides for the potential liability of platforms for failing to block and prevent certain content from being posted or shared by users. If a user’s content qualifies as materially deceptive, and other conditions are met, then the platform can be held liable for it.

Supporters point to the fact that monetary damages are not available and injunctive relief is essentially the only remedy available. The bill does allow for attorneys’ fees and costs, which could be considered the type of liability that triggers Section 230’s preemptive effect. The author has agreed to amendments that remove these remedies, leaving only injunctive relief. While courts, including the California Supreme Court, have found Section 230 immunity can extend to liability for solely injunctive relief, it is far from settled law in the country.¹⁷

In addition, the bill provide that if the platform engages in content moderation that restricts access to a candidate’s deceptive portrayal of themselves (with the required disclosure and during the applicable time period), the platform can be held liable for that content moderation decision, regardless of the justification. As discussed below, the author has agreed to an amendment that removes this provision.

Ultimately, the bill is likely to face challenge on these grounds but these amendments work toward insulating the bill from such a challenge.

¹³ *Id.*, § 230(c)(1).

¹⁴ *Id.*, § 230(c)(1) & (2).

¹⁵ *Id.*, § 230(e)(1) & (3).

¹⁶ Kosseff, *supra*, fn. 13, at pp. 94-95; see, e.g., *Doe v. MySpace Inc.* (5th Cir. 2008) 528 F.3d 413, 421-422; *Carfano v. Metrosplash.com, Inc.* (9th Cir. 2003) 339 F.3d 1119, 1125; *Zeran v. America Online, Inc.* (4th Cir. 1997) 129 F.3d 327, 333-334.

¹⁷ *Hassell v. Bird* (2018) 5 Cal. 5th 522, 547.

b. First Amendment

The First Amendment, as applied to the states through the Fourteenth Amendment, prohibits Congress or the states from passing any law “abridging the freedom of speech.”¹⁸ “[A]s a general matter, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content.”¹⁹ However, while the amendment is written in absolute terms, the courts have created a handful of narrow exceptions to the First Amendment’s protections, including “true threats,”²⁰ “fighting words,”²¹ incitement to imminent lawless action,²² defamation,²³ and obscenity.²⁴ Moreover, the First Amendment not only protects the right to speak, as a logical corollary it protects the “right to receive information and ideas.”²⁵ Expression on the internet is given the same measure of protection granted to in-person speech or statements published in a physical medium.²⁶

“Laws that burden political speech are subject to strict scrutiny, which requires the Government to prove that the restriction furthers a compelling interest and is narrowly tailored to achieve that interest.”²⁷ Content-based restrictions subject to strict scrutiny are “presumptively unconstitutional.”²⁸ California courts have been clear that political expression in the context of campaigns of any manner should be given wide latitude:

Hyperbole, distortion, invective, and tirades are as much a part of American politics as kissing babies and distributing bumper stickers and pot holders. Political mischief has been part of the American political scene since, at least, 1800.

In any election, public calumny of candidates is all too common. “Once an individual decides to enter the political wars, he subjects himself to this kind of treatment. . . . [D]eeply ingrained in our political history is a tradition of free-wheeling, irresponsible, bare knuckled, Pier 6, political brawls.” To endure the animadversion, brickbats and skullduggery of a given campaign, a politician must be possessed with the skin of a

¹⁸ U.S. Const., 1st & 14th amends.

¹⁹ *Ashcroft v. American Civil Liberties Union* (2002) 535 U.S. 564, 573.

²⁰ *Snyder v. Phelps* (2011) 562 U.S. 443, 452.

²¹ *Cohen v. California* (1971) 403 U.S. 15, 20.

²² *Virginia v. Black* (2003) 538 U.S. 343, 359.

²³ *R.A.V. v. St. Paul* (1992) 505 U.S. 377, 383.

²⁴ *Ibid.*

²⁵ *Stanley v Georgia* (1969) 394 U.S. 557, 564. Internal citations omitted

²⁶ *Reno v. ACLU* (1997) 521 U.S. 844, 870.

²⁷ *Citizens United v. FEC* (2010) 558 U.S. 310, 340. Internal citations omitted. It should be noted that while not controversial for the principle cited herein, this opinion is widely criticized for further tilting political influence toward wealthy donors and corporations.

²⁸ *Reed v. Town of Gilbert* (2015) 135 S.Ct. 2218, 2226 (*Reed*).

rhinoceros. Harry Truman cautioned would-be solons with sage advice about the heat in the kitchen.

Nevertheless, political campaigns are one of the most exhilarating phenomena of our democracy. They bring out the best and the worst in us. They allow candidates and their supporters to express the most noble and, lamentably, the most vile sentiments. They can be fractious and unruly, but what they yield is invaluable: an opportunity to criticize and comment upon government and the issues of the day.

The candidate who finds himself or herself the victim of misconduct is not without a remedy. Those campaign tactics which go beyond the pale are sanctionable under FPPC laws.

It is abhorrent that many political campaigns are mean-spirited affairs that shower the voters with invective instead of insight. The elimination from political campaigns of opprobrium, deception and exaggeration would shed more light on the substantive issues, resulting in a more informed electorate. It would encourage more able people to seek public office. But to ensure the preservation of a citizen's right of free expression, we must allow wide latitude.²⁹

The United States Supreme Court has emphasized the extraordinary protection afforded to political speech:

Discussion of public issues and debate on the qualifications of candidates are integral to the operation of the system of government established by our Constitution. The First Amendment affords the broadest protection to such political expression in order "to assure [the] unfettered interchange of ideas for the bringing about of political and social changes desired by the people." Although First Amendment protections are not confined to "the exposition of ideas," "there is practically universal agreement that a major purpose of that Amendment was to protect the free discussion of governmental affairs,... of course includ[ing] discussions of candidates...." This no more than reflects our "profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open." In a republic where the people are sovereign, the ability of the citizenry to make informed choices among candidates for office is essential, for the identities of those who are elected will inevitably shape the course that we follow as a nation. As the Court observed in *Monitor Patriot Co. v. Roy*, 401 U.S. 265, 272 (1971), "it can hardly be doubted that the constitutional guarantee has its fullest and

²⁹ *Beilenson v. Superior Court* (1996) 44 Cal. App. 4th 944, 954-55. Internal citations omitted.

most urgent application precisely to the conduct of campaigns for political office.”³⁰

This protection does not end where the truth of the speech does. “Although false statements of fact, by themselves, have no constitutional value, constitutional protection is not withheld from all such statements.”³¹ For instance, in the seminal opinion in *New York Times Co. v. Sullivan* (1964) 376 U.S. 254, 279-80, the court found the Constitution requires a rule that “prohibits a public official from recovering damages for a defamatory falsehood relating to his official conduct unless he proves that the statement was made ‘with actual malice’ -- that is, with knowledge that it was false or with reckless disregard of whether it was false or not. The Supreme Court has expounded on this principle, providing nuance based on the knowledge of the speaker:

Truth may not be the subject of either civil or criminal sanctions where discussion of public affairs is concerned. And since “. . . erroneous statement is inevitable in free debate, and . . . it must be protected if the freedoms of expression are to have the ‘breathing space’ that they ‘need . . . to survive’ . . . ,” only those false statements made with the high degree of awareness of their probable falsity demanded by *New York Times* may be the subject of either civil or criminal sanctions. For speech concerning public affairs is more than self-expression; it is the essence of self-government. The First and Fourteenth Amendments embody our “profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open, and that it may well include vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials.”

The use of calculated falsehood, however, would put a different cast on the constitutional question. Although honest utterance, even if inaccurate, may further the fruitful exercise of the right of free speech, it does not follow that the lie, knowingly and deliberately published about a public official, should enjoy a like immunity. At the time the First Amendment was adopted, as today, there were those unscrupulous enough and skillful enough to use the deliberate or reckless falsehood as an effective political tool to unseat the public servant or even topple an administration. That speech is used as a tool for political ends does not automatically bring it under the protective mantle of the Constitution. For the use of the known lie as a tool is at once at odds with the premises of democratic government and with the orderly manner in which economic, social, or political change is to be effected. Calculated falsehood falls into that class of utterances which “are no essential part of any exposition of ideas, and are

³⁰ *Buckley v. Valeo* (1976) 424 U.S. 1, 14-15. Internal citations omitted.

³¹ *People v. Stanistreet* (2002) 29 Cal. 4th 497, 505.

of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality. . . .” Hence the knowingly false statement and the false statement made with reckless disregard of the truth, do not enjoy constitutional protection.³²

As stated, a restriction can survive strict scrutiny only if it uses the least-restrictive means available to achieve a compelling government purpose.³³ This bill implicates both the right to speak about elections, as well as the right to receive information regarding them. The bill is aimed at protecting the integrity of our elections, arguably a clearly compelling governmental interest. The question is whether the bill sufficiently tailors its provisions to effectuating that goal.

The bill seeks to prevent “materially deceptive content,” which is audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media, when it portrays candidates or elections officials doing or saying something they did not do or say. However, it does not target the person creating, posting, or sharing such content, but the platforms that host it. The bill attempts to tailor itself to the boundaries sketched out above. For instance, it imposes liability on the platforms only where they knew or should have known the content qualified as “materially deceptive content.” However, this falls short of the malice standard set forth in *Sullivan*, establishing something akin to a negligence standard instead.

The bill does impose a malice requirement but on the person or entity who created the content, requiring that they created it knowing it was false or with reckless disregard for the truth. However, liability is not imposed on the creator, nor even the one posting or sharing the content, but the social media platform allowing it on their platform. Further undercutting this element, there is a rebuttable presumption that the person who created it acted with malice if the content causes “a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.” Therefore, the bill puts the onus on the platform to establish that the creator of the content, a person or entity the platform may not even have a relationship with or know, did not act with malice. Many of the relevant cases stress that the level of burden placed on a defendant to defend their political speech is a factor to consider. For instance, the following was stated in *Sullivan*:

A rule compelling the critic of official conduct to guarantee the truth of all his factual assertions -- and to do so on pain of libel judgments virtually

³² *Garrison v. Louisiana* (1964) 379 U.S. 64, 74-75. Internal citations omitted.

³³ *United States v. Playboy Entertainment Group* (2000) 529 U.S. 803, 813.

unlimited in amount -- leads to a comparable “self-censorship.” Allowance of the defense of truth, with the burden of proving it on the defendant, does not mean that only false speech will be deterred. Even courts accepting this defense as an adequate safeguard have recognized the difficulties of adducing legal proofs that the alleged libel was true in all its factual particulars.³⁴

While the plaintiff is required to prove their case by clear and convincing evidence, the standards above place a burden on platforms to establish facts potentially well outside their bounds of knowing.

The California Initiative for Technology & Democracy (CITED), the sponsor of the bill, argues the case:

AB 2655's approach is narrowly tailored and does not extend the law to hot button controversies or inflammatory claims – it does not ask social media platforms to adjudicate controversial opinions post by post. It simply stops the use of obviously, demonstrably untrue and provably false content meant to impermissibly influence our elections at peak election times. It is therefore respectful of the protections of the First Amendment and avoids concerns based on Section 230 of the Communications Decency Act.

Writing in opposition, ACLU California Action assesses the issue:

Digitally modified content or content created using artificial intelligence (AI) tools is also entitled to [First Amendment] protections, unless the content falls within recognized First Amendment exceptions such as libel or fraud. The “novelty of deepfake technology and the speed with which it is improving” do not justify relaxing the stringent protections afforded to political speech by the First Amendment. The Supreme Court has held that “whatever the challenges of applying the Constitution to ever-advancing technology, ‘the basic principles of freedom of speech and the press, like the First Amendment’s command, do not vary’ when a new and different medium for communication appears.”

The law has long made clear that the First Amendment was intended to create a wide berth for political speech because it is the core of our democracy. The First Amendment provides robust protection for speech of all kinds. Speech that is false, confusing, or which presents content that some find abhorrent, nevertheless maintains its constitutional protections as a driver of free discourse. This remains so no matter what the

³⁴ *N.Y. Times Co. v. Sullivan*, at 279.

technology used to speak. Unfortunately, the provisions of AB 2655 as currently drafted threaten to intrude on those rights and deter that vital speech.

In response to these concerns, the author has agreed to amendments that remove the provision that applies this malice standard to the creators of the content and instead more closely hews the platform's basis for liability to the malice standard, holding the large online platform liable only if it knows that the materially deceptive content meets the requirements of the bill or acts with a reckless disregard for the truth.³⁵

As the bill also requires platforms to allow certain potentially misleading content to be posted, the bill could be found to implicate the First Amendment rights of platforms in their editorial discretion. Two laws in Florida and Texas that similarly seek to prevent platforms from taking down certain content have been challenged. The consolidated case has been argued before the United States Supreme Court and an opinion is forthcoming. The 11th Circuit Court of Appeals laid out its assessment of the First Amendment implications of such laws:

Social-media platforms like Facebook, Twitter, YouTube, and TikTok are private companies with First Amendment rights and when they (like other entities) "disclos[e]," "publish[]," or "disseminat[e]" information, they engage in "speech within the meaning of the First Amendment." More particularly, when a platform removes or deprioritizes a user or post, it makes a judgment about whether and to what extent it will publish information to its users—a judgment rooted in the platform's own views about the sorts of content and viewpoints that are valuable and appropriate for dissemination on its site. As the officials who sponsored and signed S.B. 7072 [the challenged Florida law] recognized when alleging that "Big Tech" companies harbor a "leftist" bias against "conservative" perspectives, the companies that operate social-media platforms express themselves (for better or worse) through their content-moderation decisions. When a platform selectively removes what it perceives to be incendiary political rhetoric, pornographic content, or public-health misinformation, it conveys a message and thereby engages in "speech" within the meaning of the First Amendment.

Laws that restrict platforms' ability to speak through content moderation therefore trigger First Amendment scrutiny.³⁶

³⁵ This amendment includes corresponding changes in the labeling section of the bill.

³⁶ *NetChoice, LLC v. AG, Fla.* (11th Cir. 2022) 34 F.4th 1196, 1210. Internal citations and quotations omitted.

As constitutional analysis is subject to changing norms and interpretations, especially in the more political charged federal judiciary of the day, it is inherently difficult to predict whether this law will be struck down for violating the protections of the First Amendment. However, it is safe to say it will likely face legal challenge and arguably be vulnerable thereto.

In order to insulate the bill from such challenge, the author has agreed to an amendment that simply provides that the bill does not apply to a candidate's portrayal of themselves doing or saying something that the candidate did not do or say, where it includes the required disclosure.

c. Additional concerns

The bill raises a few additional concerns. First, the bill requires platforms to retain all content they have prevented or blocked or labeled pursuant to the bill. This forced retention of information raises some thorny legal issues and may interfere with existing consumer rights. For instance, the CCPA, as amended by the CPRA, grants a series of rights to consumers, including the right to delete information held by businesses. In addition, given that the retention provision is essentially a government mandate on private businesses to seize certain information of private individuals, Fourth Amendment issues arguably arise. Furthermore, the bill requires platforms to hand over the content to specified government entities and even "researchers," upon request. There is no limitation that there be evidence of a crime or some other justification and no probable cause necessary to be provided the information. In response, the author has agreed to an amendment to remove this retention requirement.

In addition, it is unclear what exactly is required by the bill's requirement to block or prevent the *sending* of materially deceptive content. This could be read to apply to private messaging features of these platforms, essentially requiring platforms to scan private communications. This would raise serious privacy concerns. In response, the author has agreed to amendments that remove the "sending" element of the bill.

In addition, groups in opposition raise concerns that the bill presupposes a level of sophistication for technology that can detect AI-generated or manipulated content that simply does not exist. A coalition of industry associations, including NetChoice writes in opposition:

AB 2655 appears to be based on the false assumption that online platforms definitively know whether any particular piece of content has been manipulated in such a way that is defined under the bill. While digital services may employ tools to identify and detect these materials with some degree of certainty, it is an evolving and imperfect science in its current form. AB 2655 also presumes that online platforms are an appropriate arbiter of deciding what constitutes accurate election

information. However, most digital services are not equipped with the tools or expertise to make such judgments.

Oakland Privacy writes in opposition:

The bill language offers that a technology company should be the judge, jury and executioner, although it may be unclear if the content is or is not generative AI-created and what role generative AI played in the content. It is unclear to us how any technology platform can be expected to know everything that every candidate in every city, county, state and federal election said and everywhere they went. Not to mention every other elected official in the state. If this is the basis for the removal of content by a technology platform, it is highly speculative and largely dependent on reports to the platform, which may be inaccurate, politically motivated, or malicious.

We appreciate amendments to raise the bar for the knowledge level of online platforms. But we continue to have concerns on the other side of the spectrum: the removal of content that should not be removed and may well impact election results.

In other words, the bill language is relying on two imprecise measures: technically scanning content for synthetic material with highly inaccurate tools, and real-life reports from the public, candidates and election officials and campaigns or chaos actors to power a broad censorship regime of blocking content. We cannot support that, even under the guise of defending democracy.

The opposition coalition also takes issue with the enforcement mechanism:

[B]ecause AB 2655 is focused on enforcement against covered platforms and not the actors who are intentionally seeking to materially deceive other consumers, it is unlikely to meaningfully reduce the amount of election mis- and disinformation hosted online. While the June 11 amendments appear to attempt to address this issue, we do not believe the new language effectively resolves our concerns. For example, the bill now allows for a "rebuttable presumption" but still fails to effectively address and hold accountable the purveyors of deceptive content.

4. Support

CITED, the sponsor of the bill, writes:

Those trying to influence campaigns – conspiracy theorists, foreign states, online trolls, and candidates themselves – are already creating and

distributing election-threatening deepfake images, audio, and video content in the US and around the world. This threat is not imaginary: generative AI has been used in various ways – most of them deeply deceptive – to influence the national elections in Slovakia, Bangladesh, Argentina, Pakistan, and elsewhere, including in our own country. Examples of this occurring in U.S. elections include Ron Desantis using AI-generated images to attack his opponent in his presidential run, foreign states caught attempting to influence American politics through social media, and just this month, a supporter of former President Trump creating a deepfake image of Trump with Black Americans designed to persuade Black voters to support Trump.

These examples demonstrate the power of generative AI-fueled disinformation to skew election results and weaken our faith in our democracy. We cannot let it undermine our elections here in California, and we are grateful you are leading the effort to try to stop it.

AB 2655 strikes the right balance by seeking to ban, for a strictly limited time before and after elections, the online spread of the worst deepfakes and disinformation maliciously intended to prevent voters from voting or getting them to vote erroneously based on fraudulent content. The bill also requires that other fake online content related to elections and elections processes (such as redistricting), which is also designed to undermine election procedures and democratic institutions, must be labeled as fake, again just for a limited time. The bill only applies to the largest online platforms with the greatest reach of potential election disinformation, and we believe it is fully implementable today based on tools these companies already possess. The companies covered by the bill's requirements are all already subject to similar requirements under the European Union's Digital Services Act, which is designed to, among other things, crack down on election interference.

A coalition of groups in support, including the American Federation of State, County, and Municipal Employees (AFSCME) and NextGen CA, write:

AB 2655 seeks to solve these problems by, for a limited time before and after elections, banning the online spread of the worst of the deepfakes and disinformation meant to prevent voters from voting or to deceive them based on fraudulent content, and requiring that other fake content to be labeled as such. The approach leans heavily on increasing transparency, with bans used at only the highest-leverage moments, making it narrowly tailored. Additionally, it does not extend the law to hot button controversies or inflammatory claims – just the depiction of demonstrably untrue and provably false content meant to impermissibly

influence our elections, at peak times – and is therefore implementable and respectful of the protections of the First Amendment.

Writing in support, the Northern California Recycling Association explains the need for the bill:

California is entering its first-ever generative Artificial Intelligence (AI) election, in which disinformation powered by generative AI would and will pollute our information ecosystems like never before. Deepfakes are a powerful and dangerous tool in the arsenal of those that want to wage disinformation campaigns, and they have the potential to wreak havoc on our democracy by attributing speech and conduct to a person that is false or that never happened. Advances in AI make it easy for practically anyone to generate this deceptive content, making it that much more important that we identify and restrict its spread before it has the chance to deceive voters and undermine our democracy.

SUPPORT

California Initiative on Technology and Democracy (sponsor)
AFSCME California
Asian Americans Advancing Justice - Asian Law Caucus
Asian Americans and Pacific Islanders for Civic Empowerment
Asian Law Alliance
Bay Rising
Board of Supervisors for the City and County of San Francisco
California Clean Money Campaign
California Environmental Voters
California State Sheriff's Association
California Voter Foundation
Center for Countering Digital Hate
Chinese Progressive Association
City and County of San Francisco Board of Supervisors
Courage California
Disability Rights California
Hmong Innovating Politics
Inland Empire United
League of Women Voters of California
Move (mobilize, Organize, Vote, Empower) the Valley
Nextgen California
Northern California Recycling Association
Partnership for the Advancement of New Americans
SEIU California
Techequity Action

Verified Voting
Young People's Alliance
Youth Power Project

OPPOSITION

ACLU California Action
California Chamber of Commerce
Computer & Communications Industry Association
Electronic Frontier Foundation
Internet Works
Netchoice
Oakland Privacy
Software & Information Industry Association
Technet

RELATED LEGISLATION

Pending Legislation:

SB 942 (Becker, 2024) establishes the California AI Transparency Act, requiring covered providers to create and make freely available an AI detection tool to detect content as AI-generated and to include disclosures in content generated by the provider's system. SB 942 is currently in the Assembly Judiciary Committee.

SB 970 (Ashby, 2024) ensures that media manipulated or generated by artificial intelligence (AI) technology is incorporated into the right of publicity law and criminal false impersonation statutes. The bill requires those providing access to such technology to provide a warning to consumers about liability for misuse. SB 970 was held on suspense in the Senate Appropriations Committee.

AB 2355 (Wendy Carrillo, 2024) requires committees that create, publish, or distribute a political advertisement that contains any image, audio, or video that is generated or substantially altered using artificial intelligence to include a disclosure in the advertisement disclosing that the content has been so altered. AB 2355 is currently in this Committee.

AB 2839 (Pellerin, 2024) prohibits a person, committee, or other entity from knowingly distributing an advertisement or other election communication that contains materially deceptive content, as defined and specified, with malice, except as provided, within 120 days of a California election, and in specified cases, 60 days thereafter. AB 2839 is currently in this Committee.

AB 2930 (Bauer-Kahan, 2024) requires, among other things, a deployer and a developer of an automated decision tool to perform an impact assessment for any automated

decision tool the deployer uses that includes, among other things, a statement of the purpose of the automated decision tool and its intended benefits, uses, and deployment contexts. AB 2930 requires a deployer to, at or before the time an automated decision tool is used to make a consequential decision, notify any natural person that is the subject of the consequential decision that an automated decision tool is being used to make, or be a substantial factor in making, the consequential decision and to provide that person with, among other things, a statement of the purpose of the automated decision tool. AB 2930 is currently in this Committee.

AB 3211 (Wicks, 2024) establishes the California Provenance, Authenticity and Watermarking Standards Act, which requires a generative AI system provider to take certain actions to assist in the disclosure of provenance data to mitigate harms caused by inauthentic content, including placing imperceptible and maximally indelible watermarks containing provenance data into content created by an AI system that the generative AI system provider makes available. AB 3211 also requires a large online platform, as defined, to, among other things, use labels to prominently disclose the provenance data found in watermarks or digital signatures in content distributed to users on its platforms, as specified. AB 3211 is currently in the Senate Appropriations Committee.

Prior Legislation: AB 730 (Berman, Ch. 493, Stats. 2019) prohibited the use of deepfakes depicting a candidate for office within 60 days of the election unless the deepfake is accompanied by a prominent notice that the content of the audio, video, or image has been manipulated. Additionally, AB 730 authorized a candidate who was falsely depicted in a deepfake to seek rapid injunctive relief against further publication and distribution of the deepfake.

PRIOR VOTES:

Senate Elections and Constitutional Amendments Committee (Ayes 6, Noes 1)

Assembly Floor (Ayes 56, Noes 1)

Assembly Appropriations Committee (Ayes 11, Noes 1)

Assembly Judiciary Committee (Ayes 9, Noes 0)

Assembly Elections Committee (Ayes 6, Noes 1)

Exhibit 4

Date of Hearing: April 23, 2024

ASSEMBLY COMMITTEE ON JUDICIARY
Ash Kalra, Chair
AB 2655 (Berman) – As Amended April 1, 2024

As Proposed to be Amended

SUBJECT: DEFENDING DEMOCRACY FROM DEEPFAKE DECEPTION ACT OF 2024

KEY ISSUE: SHOULD LARGE ONLINE PLATFORMS BE REQUIRED TO BLOCK (OR IN SOME CASES LABEL) MATERIALLY DECEPTIVE AND DIGITALLY MODIFIED OR CREATED CONTENT RELATED TO AN ELECTION OR ELECTION PROCESS, DURING A PRESCRIBED PERIOD OF TIME BEFORE OR AFTER AN ELECTION?

SYNOPSIS

According to the author, disinformation powered by Artificial Intelligence (AI), which can be distributed to millions of social media users in an instant, poses a serious threat to our political discourse, our elections, and indeed our democracy. For example, “deepfakes” can generate false sounds and images that could lead to the most discerning viewer to falsely conclude that a candidate, elected official, or election worker said or did something they did not do. Such disinformation not only distorts the truth, it has the potential to undermine people’s confidence in our political institutions. While disinformation can threaten political discourse at any time, the author believes that it is especially harmful during an election season, when uncorrected disinformation may influence an election result or create false concerns about the legitimacy of an election in its immediate aftermath.

This bill, therefore, would require a large online platform to block “materially deceptive and digitally modified or created” content that portrays any candidate, elected official, or elections official doing or saying something they did not do or say; it would also require them to block deceptive material that concerns voting machines, ballots, voting sites, or other procedures or equipment related to an election. In addition, deceptive material about broader “elections processes,” that are not subject to blocking requirement, would need to contain a label that they are materially deceptive and digitally modified. The bill would allow any resident of California to inform the platform that covered material had not been properly blocked or labeled, and if the platform does not respond within 36 hours, or if the reporting resident does not agree with the response, the resident may bring an action for injunctive relief. The bill would also allow the Attorney General, or any district attorney or city attorney, to also seek injunctive relief.

This bill passed out of the Assembly Elections Committee on a 6-1 vote. It is sponsored by the California Initiative for Technology and Democracy, a project of California Common Cause, and supported by several political reform groups and labor organizations, among others. The bill is opposed by groups representing the information and technology industry and by ACLU Action California. The opposition argues that the bill would be ineffective, unconstitutional, and preempted by federal law. The author will take several definitional amendments in this Committee, which are reflected in the Summary, below, and discussed in the analysis.

SUMMARY: Requires large online platforms, as defined, to block the posting or sending of materially deceptive and digitally modified or created content related to elections, or to label that content, during specified periods before and after an election. Specifically, **this bill**:

- 1) Makes findings and declarations about the growing use of generative artificial intelligence (AI), deepfakes, and related technologies to disseminate disinformation that distorts our electoral process and undermines trust in elections.
- 2) Requires a large online platform (platform), using state-of-the-art, best available tools to detect digitally modified or created content, to develop procedures for blocking and preventing the posting or sending of materially deceptive and digitally modified or created content, and to block and prevent that content if the platform knows or should know that the content meets the following requirements, during a specified time period, of any of the following:
 - a) A candidate portrayed as doing or saying something that the candidate did not do or say.
 - b) An elections official portrayed as doing or saying something in connection with the performance of their elections-related duties that the official did not do or say.
 - c) An elected official portrayed as doing or saying something that influences the election that the elected official did not do or say.
 - d) A voting machine, ballot, voting site, or other property or equipment related to an election that is portrayed in a materially false way.
- 3) Prohibits a platform, notwithstanding the above, from preventing candidates from posting deceptive and manipulated material *about themselves* so long as the content includes a prescribed disclaimer.
- 4) Provides that the above prohibitions apply only during the period 120 days before and through election day; however, they apply to content relating to elections officials, voting sites, voting machines, ballots, or related equipment from 120 days before the election to 60 days after the election.
- 5) Requires a platform to develop procedures for labeling materially deceptive and digitally modified or created content that pertains to election processes, but that is not subject to the blocking provisions above. Requires the label to indicate that the content is inauthentic, fake, or false if the platform knows or should know as much. Provides that the labeling requirement applies during the period beginning one year before the election or election process, as specified.
- 6) Requires the platform to provide a way for Californians to report content that was not blocked or labeled as required, and requires the platform to respond to the report within 36 hours and to describe any actions taken. If the platform does not respond within 36 hours, or if the reporting resident disagrees with the response, the reporting resident may bring an action for injunctive or other equitable relief, and a prevailing plaintiff may recover reasonable attorney's fees and costs. Allows the Attorney General, a district attorney, or a city attorney to similarly seek injunctive relief and obtain fees and costs.

- 7) Specifies that the provisions of this bill do not apply to a regularly published online newspaper, magazine, or periodical, as specified.
- 8) Defines the following terms for purposes of the above as follows:
 - a) “Elections official” means any of the following: an elections official as defined in Elections Code Section 320; the Secretary of State and their staff; a temporary worker, poll worker, or member of a precinct board; any other person charged with holding or conducting an election, a canvas, or performing another election-related duty.
 - b) “Election processes” means any government process related to an election, including, but not limited to, elections, candidates, vote counting, redistricting, and proceedings or processes of the Electoral College.
 - c) “Materially deceptive and digitally modified or created content” means an image or an audio or video recording or other digital content, including a chatbot, that has been intentionally manipulated such that all of the following conditions are met:
 - i) The digital content is the product of digital manipulation, artificial intelligence, or machine learning, including deep learning techniques, that merges, combines, replaces, or superimposes content onto an image or an audio or video recording, creating an image or an audio or video recording that appears authentic, or that otherwise generates an inauthentic image or an audio or video recording that appears authentic, and that contains a false portrayal of any of the following: a candidate for elective office, elected official, elections official, voting machine, ballot, voting site, other property or equipment related to an election, or elections process; and provides that “false portrayal” means the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if they were hearing or seeing the unaltered, original version of the content.
 - ii) The person or entity who attempted to post or send, or who did post or send, the content did so knowing the portrayal was false, or did so with reckless disregard for whether the portrayal was false. If the content is intentionally manipulated and contains a false portrayal as specified in subparagraph (A), there shall be a rebuttable presumption that the person or entity knew the portrayal was false or that they acted with reckless disregard for whether the portrayal was false.
 - d) Clarifies that “Materially deceptive and digitally modified or created content” does not include any image or audio or video recording that contains only minor modifications that do not lead to significant changes to the perceived contents or meaning of the content. Minor changes include changes to the brightness or contrast of images, removal of background noise in audio, and other minor changes that do not impact the content of the image or audio or video recording.
 - e) “Large online platform” means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1 million California users during the preceding 12 months.

- f) “Artificial intelligence” means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

EXISTING LAW:

- 1) Prohibits a person, committee, or other entity, until January 1, 2027, from distributing with actual malice, within 60 days of an election at which a candidate for elective office will appear on the ballot, materially deceptive audio or visual media of a candidate with the intent to injure the candidate’s reputation or to deceive a voter into voting for or against the candidate.
 - a) Defines “materially deceptive audio or visual media,” for these purposes, as an image or an audio or visual recording of a candidate’s appearance, speech or conduct that has been intentionally manipulated in a manner that both of the following are true about the image or audio or video recording: (1) It would falsely appear to a reasonable person to be authentic; and (2) it would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than the person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording.
 - b) Provides that this prohibition does not apply if the audio or visual media includes a disclaimer stating “This (image/video/audio) has been manipulated,” and the disclaimer complies with specified requirements.
 - c) Permits a candidate whose voice or likeness appears in deceptive audio or visual media distributed in violation of this provision to seek injunctive relief, as specified, and general or special damages and reasonable attorney’s fees and costs, as specified. Specifies that in any civil action brought pursuant to these provisions, the plaintiff bears the burden of establishing the violation through clear and convincing evidence.
 - d) Provides that this prohibition shall not be construed to alter or negate any rights, obligations, or immunities of an interactive service provider under Section 230 of the federal Communications Decency Act.
 - e) Provides that this prohibition does not apply to a radio or television broadcasting station; an internet website; a regularly published newspaper, magazine, or other periodical of general circulation, including an internet or electronic publication; or media that constitute satire or parody. (Elections Code Section 20010.)
- 2) Prohibits a person, firm, association, corporation, campaign committee, or organization, beginning January 1, 2027, with actual malice, from producing, distributing, publishing, or broadcasting campaign material, as defined, that contains either of the following types of pictures or photographs, as specified, unless the campaign material includes a disclaimer that the picture is not an accurate representation of fact:
 - a) A picture or photograph of a person or persons into which the image of a candidate for public office is superimposed.

- b) A picture or photograph of a candidate for public office into which the image of another person or persons is superimposed. (Elections Code Section 20010.)

FISCAL EFFECT: As currently in print this bill is keyed fiscal.

COMMENTS: According to the author:

AB 2655 will ensure that online platforms restrict the spread of election-related deceptive deepfakes meant to prevent voters from voting or to deceive them based on fraudulent content. Deepfakes are a powerful and dangerous tool in the arsenal of those that want to wage disinformation campaigns, and they have the potential to wreak havoc on our democracy by attributing speech and conduct to a person that is false or that never happened. Advances in AI make it easy for practically anyone to generate this deceptive content, making it that much more important that we identify and restrict its spread before it has the chance to deceive voters and undermine our democracy.

Therefore, in order to ensure California elections are free and fair, online platforms must prevent the online spread of election-related deceptive deepfakes and disinformation meant to prevent voters from voting or to deceive them based on fraudulent content.

Existing laws maintaining “election integrity.” As aptly noted in the analysis of this bill by the Assembly Elections Committee, the “use of false and deceptive information in campaigns to influence election outcomes is not a new phenomenon,” nor are the laws “aimed at curbing such practices” new. Indeed, in 1850, the First Session of the California State Legislature created penalties for “deceiving [an elector] and causing him to vote for a different person for any office than such elector desired or intended to vote for.” (Chapter 38, Statutes of 1850.) Existing California law has greatly elaborated on these initial legislative efforts. For example, provisions in the Elections Code prohibit the distribution of false and misleading information about qualifications to vote or about the days, dates, times, and places where voting may occur; prohibit the misleading use of government seals in campaign literature (Elections Code Section 18304); and prohibit coercing or deceiving people into voting in a way that is inconsistent with the person’s intent (Elections Code Sections 18302, 18304, 18573 and 18573.5).

In the last five years, the Legislature has turned its attention to “materially deceptive” audio and visual materials that portray a candidate. For example, AB 730 (Chap. 493, Stats. 2019) responded to reports that so-called “deepfake” technology, a software that allows someone to produce audios and videos that look and appear remarkably real to even the most discerning person. For example, in 2018, BuzzFeed and the film director Jordan Peele published a very realistic-looking deepfake showing former President Obama calling then-President Donald Trump a “total and complete dipshit.” Obama did not say that. Peele and BuzzFeed did not use the video to try to influence a political election – indeed halfway through the video the ruse was revealed – but to show the potential for abuse of deepfake technology. Responding to this and similar reports, AB 730 prohibited the distribution of materially deceptive audio or visual media with the intent to injure the candidate’s reputation or to deceive a voter into voting for or against a candidate.

AB 730 was itself an amendment to California’s “Truth in Political Advertising Act” of 1998, which had prohibited campaign material that deceptively altered a picture of a candidate (for example by superimposing one person’s image upon another) unless the picture contained a disclaimer. This 1998 bill was introduced in response to the use of technologies like “photo

shopping,” which now seem quaint compared to deepfakes, including deepfakes generated by artificial intelligence. The bill now before the Committee, like AB 730 before it, is apparently an effort to stay one step ahead of evolving technologies, which not only create more realistic-looking deceptions, but make it possible to quickly create, alter, and distribute fake images to millions of people in the blink of an eye (or the click of a mouse).

This bill and existing law: who? AB 2655 elaborates upon and modifies existing law in a variety of ways. Most significant, existing law prohibits *a person, committee, or other entity* from distributing, with actual malice, “materially deceptive” audio or visual media of a candidate with the intent to injure the candidate’s reputation or to deceive a voter into voting for or against the candidate. Existing law prohibits the distribution of such material within 60 days of the election in which the targeted candidate is running for office. The prohibitions in AB 2655, on the other hand, do not apply to the person or entity that created the materially deceptive material, but to the “large online platform” on which it is posted. AB 2655 requires the online platform to develop and implement procedures to block or prevent the posting of the covered content.

This bill and existing law: what? This bill also differs from existing law in terms of covered content. Existing law applies only to manipulated material that misrepresents the “candidate” saying or doing something the candidate did not do or say. This bill similarly applies to content deceptively depicting the candidate, but also applies to images portraying an “elections official” doing or saying something they did not do or say, as well as images depicting a voting machine, ballot, voting site, or voting equipment in a materially false way. In addition, this bill would require the platform to label (but not necessarily block) materially deceptive content about “election processes” (as opposed to a specific candidate, election official, or voting site). Although the distinction between what material must be blocked versus what material must be labeled is not entirely clear, the intent of the author and sponsor is that the more the material singles out a particular candidate or election official, the more it must be blocked; whereas material that deals with “election processes” more generally need only be labeled. *The author has agreed to work with the Committee as the bill moves forward to clarify the distinction between the blocking and labeling requirements.*

This bill and existing law: when? AB 2655 also expands and modifies the relevant time period that prohibitions are in force. Existing law prohibits someone from distributing deceptive material during the 60-day period before an election. This bill, however, requires the platform to block covered material a period beginning 120 days before the election and through the day of the election. However, if the deceptive material pertains to an election official, or deceptively depicts a voting machine, ballot, voting site, or property or equipment related to an election, the platform must also block the content for 60 days *after* the election. Presumably, this post-election period is intended to prevent false claims – similar to those made in 2020 – that the election process was irregular or otherwise “rigged.” The labeling requirement covers an even larger time frame; it applies during the period beginning *one year* before the “election.” The bill also requires labeling for the period beginning one year before an “election process.” “Election processes” – as opposed to an “election” – is defined to include any government process “related” to an election, “including, but not limited to,” elections, candidates, vote counting, redistricting, and proceedings or processes of the electoral college.” Because an “election” has a known date, it should be fairly easy for the platform to figure out when the year-long labeling period starts. However, if the platform must also label one year prior to an “election process,” when does a process like “redistricting” start? Does it start with each new census? Does it start when the legislative body (or in some states a commission) meet to draw up new district lines?

Moreover, the definition of “election process” is not limited to the items listed, expressly stating “including, but is not limited to,” those items. *As discussed above, the author has agreed to work with the Committee as the bill moves forward to clarify the distinction between materials that the platform is required to “block” and those it is required to “label.”*

This bill and existing law: how enforced? In addition to differences as to who, what, and when, this bill also differs in how violations would be enforced. Existing law permits only the “candidate” whose voice or likeness appears in the deceptive material to bring an action for injunctive relief, general or special damages, and reasonable attorney’s fee and cost. However, under existing law the candidate bears the burden of establishing a violation by “clear and convincing evidence.” Enforcement provisions in this bill are much different. The bill allows any “California resident” to report to the platform that content was not blocked or labeled as required. If the platform does not respond within 36 hours, *or if the resident disagrees with the response*, the resident may seek injunctive relief to compel compliance and, if the resident prevails, shall be awarded reasonable attorney’s fees and costs. Thus, not only can any California resident – not just a person depicted or otherwise affected – seek injunctive relief, they apparently do not have the burden of proving a violation by clear and convincing evidence. If they “disagree” with the platform’s response, that’s enough; they can seek injunctive relief and a court would decide to issue on the likelihood of success on merits, the general rule for injunctive relief. *The author may wish to consider, as the bill moves forward, increasing the standard of proof to the higher clear and convincing evidence; and limiting who may seek relief.*

First Amendment concerns. Because this bill imposes a *government* mandate that online platforms must block expressive material based upon its content, it implicates the First Amendment. The First Amendment provides that “Congress shall make no law . . . prohibiting the freedom of speech.” As interpreted by the courts and incorporated against the states by the due process clause of the 14th Amendment, the First Amendment prevents any government entity (not just Congress) from enacting any law or adopting any policy that burdens freedom of speech. In addition, Article I, Section 2 of the California Constitution guarantees to every person the freedom to “speak, write, and publish his or her sentiments on all subjects, being responsible for the abuse of this right.” Moreover, the First Amendment not only protects the right to speak, as a logical corollary it protects the “right to receive information and ideas.” (*Stanley v Georgia* (1969) 394 U.S. 557, 564.) This bill would interfere with both the expression and reception of information based upon its content. Moreover, not only does this bill single out particular content, the content relates to political candidates and elections. This is potentially problematic because the First Amendment affords the “broadest protection” to the “discussion of public issues” and “political expression in order to assure the unfettered interchange of ideas for the bringing about of political and social changes desired by the people.” (*McIntyre v Ohio Election Commission* (1997) 514 U.S. 334.) It is difficult to imagine any content more related to “political expression” and “discussion of public issues” than content about candidates and elections. The fact that the bill restricts speech that is “materially deceptive” or “false” does not matter, for the U.S. Supreme Court has been unequivocal that the First Amendment protects even “false” speech. The remedy for false speech is more true speech, and false speech tends to call forth true speech. (*United States v Alvarez* (2012) 567 U.S. 709.)

The bill will likely meet the “compelling interest” threshold, if not the “narrowly tailored” threshold. The right to free speech is not absolute. As Justice Holmes noted in a dissenting opinion over a century ago, the First Amendment does not protect a right to falsely cry fire in a crowded theater. (*Schenck v. United States* (1919) 249 U.S. 47.) The proponents of this bill may,

not unreasonably, think that pervasive disinformation about candidates and elections, especially during an election season, is just as dangerous. In reviewing the law, the Court would apply strict scrutiny. This means that government can impose even content-based restrictions on protected speech *if* they have a “compelling government interest” *and* they use “narrowly tailored means” to achieve that interest.

Even the opponents of this bill appear to concede that maintaining “election integrity” is a “weighty” and, presumably, “compelling” government interest. Therefore, it seems likely that if this bill is enacted and subsequently challenged, a court will accept that there is a “compelling interest” but still need to consider whether the “means” are “narrowly tailored.” The opponents of this measure claim that it is *not* narrowly tailored. They contend that while covered platforms may have state-of-the-art tools that allow them to identify content that has been “digitally modified,” there is no technology to determine if content is “materially deceptive.” The bill defines “materially deceptive,” in relevant part, to mean content that is “intentionally manipulated” so that it appears “authentic” but contains a “false portrayal” of a candidate, elected official, elections official, voting machine, ballot, voting site, other property or equipment related to an election, or elections process. The purpose of narrow tailoring is to ensure that no more speech is infringed or burdened than is necessary. However, the opponents of this bill – both the industry groups and the ACLU – believe that with no sure means to determine what is “materially deceptive,” the platforms will err on the side of blocking content, thus burdening more speech than is necessary.

The findings and declarations in the bill state that the “labeling information required by this bill is narrowly tailored to provide consumers with factual information about the inauthenticity of particular images, audio, video, or text content in order to prevent consumer deception.” Tellingly, there is no similar claim about the blocking requirement being narrowly tailored. In any event, it will be a court – not the findings and declarations of the bill – that will determine whether the bill is narrowly tailored. The court may consider, for example, if there are other less restrictive and more effective means of protecting election integrity.

Section 230 concerns and “editorial discretion.” In addition to implicating the First Amendment, this bill may also be preempted by Section 230 of the federal Communications Decency Act. In relevant part, Section 230 provides two express protections for online platforms and their ability to moderate online content. First, Section 230 declares that an online platform cannot be held liable for content posted by third parties. The rationale for this immunity is premised on the idea that online platforms, unlike newspapers, do not exercise editorial discretion; rather, like telephone companies or “common carriers” they are merely a conduit for the expression of ideas by others. Those others, not the platform, are liable for any harm caused by the content. Second, somewhat in tension with this immunity, Section 230 expressly provides that online platforms are not liable if they block or remove material because they disapprove of its content. While it is important to remember that the First Amendment is distinct from Section 230, this protection flows from First Amendment principles. First, because the online platform is not a government actor, it cannot violate the First Amendment. Second, as a private actor, the platform has its own First Amendment right not to be associated with speech it finds objectionable. Some scholars have noted that the two protections in Section 230 are based on contrary premises. Immunity from liability from postings by third parties assumes that platforms *do not* exercise editorial discretion. Their right to remove content without liability, and their own free speech claims, on the other hand, assume that they *do* exercise editorial discretion. [For a concise overview of Section 230 and its intersection with the First Amendment, see Bollinger

and Stone, *Social Media, Freedom of Speech, and the Future of our Democracy* (Oxford University Press, 2022), especially pp. xxiii-xl; on efforts to reform Section 230, see pp.103-120.] Whatever the merits or demerits of Section 230 may be, it is federal law and appears to grant social media platforms the right to moderate content on their platforms and immunizes them from liability for content posted by the third party.

Cases pending before the U.S. Supreme Court. In February of this year, the U.S. Supreme Court heard arguments about two state laws that may have far-reaching consequences for both First Amendment case law and the status of Section 230. Largely in response to social media platforms barring former President Donald Trump from their platforms in the wake of the January 6 riots, both Texas and Florida enacted laws that limited the ability of social media platforms to control content on their platforms. The Florida law fines platforms if they ban a candidate for office in their state, and requires platforms to disclose information about their moderation policies. The Texas law prohibits platforms from removing content based on its “viewpoint.” Both of these laws directly challenge the provision in Section 230 that expressly allows platforms to remove content. Both laws also raise First Amendment concerns about the platforms’ right not to be associated with views with which they disagree. Both laws provide an interesting point of comparison with the bill under review: Texas and Florida *prohibit* a platform from denying access to certain people or blocking content on certain topics, while this bill would *require* the platforms to remove content. The Court is expected to issue a ruling in June of this year. How that ruling would affect this bill is unclear, especially given that this bill moves in the opposite direction of the Florida and Texas laws. However, if the Court decides in favor of the platforms – which many commentators think they will, at least in part – it might suggest that the Court believes that platforms should be given more freedom to self-moderate content without state interference. (See David McCabe, “Social media companies are bracing for Supreme Court arguments on Monday that could fundamentally alter how the platforms police their sites,” *New York Times*, February 25, 2024; and Adam Liptak, “The Supreme Court seemed skeptical on laws in Florida and Texas,” *Id.* February 26, 2024.)

Like constitutional and preemption questions, there is no obvious or certain answer as to whether this bill violates the First Amendment or Section 230. The Court may provide some insight soon enough.

Proposed Author Amendments. The author will take the following amendments to the definitions section of the bill:

(a) ***“Artificial intelligence” means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.***

(b) (1) “Elections official” means any of the following persons, but only in their capacity as a person charged with holding or conducting an election, conducting a canvass, assisting with the holding or conducting of an election or a canvas, or performing another duty related to administering the provisions of this code:

(A) An elections official as defined in Section 320.

(B) The Secretary of State and their staff.

(C) A temporary worker, poll worker, or member of a precinct board.

(D) Any other person charged with holding or conducting an election, conducting a canvass, assisting with the holding or conducting of an election or a canvas, or performing another duty related to administering the provisions of this code.

(2) The requirements of this chapter relating to content portraying an elections official apply only if the large online platform knows or should know that the person is an elections official.

(c) ~~(b)~~ “Election processes” means any government process related to an election, including, but not limited to, elections, candidates, vote counting, redistricting, and proceedings or processes of the electoral college.

(d) ~~(e)~~ (1) “Materially deceptive and digitally modified or created content” means an image or an audio or video recording or other digital content, including a chatbot, that has been intentionally manipulated such that all of the following conditions are met:

(A) (i) The digital content is the product of digital manipulation, *including, but not limited to, artificial intelligence*, ~~artificial intelligence, or machine learning, including deep learning techniques, that merges, combines, replaces, or superimposes content onto an image or an audio or video recording, creating an image or an audio or video recording that appears authentic, or that otherwise generates an inauthentic image or an audio or video recording that appears authentic, and that~~ *but* contains a false portrayal of any of the following: a candidate for elective office, elected official, elections official, voting machine, ballot, voting site, other property or equipment related to an election, or elections process.

(ii) For purposes of this subdivision, “false portrayal” means the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if they were hearing or seeing ~~the unaltered, original~~ *an authentic* version of the content.

(B) The person or entity who attempted to post or send, or who did post or send, the content did so knowing the portrayal was false, or did so with reckless disregard for whether the portrayal was false. If the content is intentionally manipulated and contains a false portrayal as specified in subparagraph (A), there shall be a rebuttable presumption that the person or entity knew the portrayal was false or that they acted with reckless disregard for whether the portrayal was false.

(2) “Materially deceptive and digitally modified or created content” does not include any image or audio or video recording that contains only minor modifications that do not lead to significant changes to the perceived contents or meaning of the content. Minor changes include changes to the brightness or contrast of images, removal of background noise in audio, and other minor changes that do not impact the content of the image or audio or video recording.

(e) ~~(d)~~ “Large online platform” means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.

ARGUMENTS IN SUPPORT: The sponsor of this bill – The California Initiative for Technology and Democracy (CITED) – writes in support of AB 2655:

California and the nation are entering the first-ever generative artificial intelligence (AI) election, in which disinformation powered by generative AI will pollute our information ecosystems like never before. . . Those trying to influence campaigns – conspiracy theorists, foreign states, online trolls, and candidates themselves – are already creating and distributing election-threatening deepfake images, audio, and video content in the US and around the world. . . Examples of this occurring in U.S. elections include Ron Desantis using AI-generated images to attack his opponent in his presidential run, foreign states caught attempting to influence American politics through social media, and just this month, a supporter of former President Trump creating a deepfake image of Trump with Black Americans designed to persuade Black voters to support Trump. These examples demonstrate the power of generative AI-fueled disinformation to skew election results and weaken our faith in our democracy.

AB 2655 strikes the right balance by seeking to ban, for a strictly limited time before and after elections, the online spread of the worst deepfakes and disinformation maliciously intended to prevent voters from voting or getting them to vote erroneously based on fraudulent content. . . AB 2655's approach is narrowly tailored and does not extend the law to hot button controversies or inflammatory claims – it does not ask social media platforms to adjudicate controversial opinions post by post. It simply stops the use of obviously, demonstrably untrue and provably false content meant to impermissibly influence our elections at peak election times.

ARGUMENTS IN OPPOSITION: Opponents of AB 2655 contend that the bill is unnecessary, unwise, unconstitutional, or some combination thereof. A coalition of groups representing the technology and information industry (industry opponents) contend that responsible digital service providers already “take aggressive steps to moderate dangerous and illegal content, consistent with their terms of service. The companies deliver on the commitments made to their user communities with a mix of automated tools and human review.” For example, industry opponents point to the several online businesses that voluntarily participate in “the Digital Trust & Safety Partnership (DTSP) to develop and implement best practices to ensure a safer and more trustworthy internet, and have recently reported on the efforts to implement these commitments.”

Industry opponents believe that AB 2655 falsely assumes that online platforms “definitively know whether any particular piece of content has been manipulated in such a way that is defined under the bill. While digital services may employ tools to identify and detect these materials with some degree of certainty, it is an evolving and imperfect science in its current form. AB 2655 also presumes that online platforms are an appropriate arbiter of deciding what constitutes accurate election information. However, most digital services are not equipped with the tools or expertise to make such judgments.”

In addition to these practical and operational concerns, industry opponents also question the effectiveness of the bill’s approach. For example, they point out that the bill only applies to the largest online platforms, specifically those with at least one million California users. Therefore the bill would not include platforms like Truth Social or Parler (which may relaunch this year) even though they are the ones that produce most of the concern. Opponents also point to the “sweeping” enforcement provisions, most notably the provision that allows “any California

resident” to notify the platform of content that, in the resident’s opinion, should have been blocked or labeled. The bill, opponents note, would allow this resident “to bring a civil action against a large online platform if the platform has not responded within 36 hours or if the reporting resident disagrees with the platform’s response.” Confronted with such a restricted timeline and the threat of a civil action, the opponents contend, platforms will “remove significantly more content, including content that has *accurate* election information and content that is not materially deceptive.”

While industry opponents concentrate on problems of implementation and effectiveness, ACLU California Action focuses on the bill’s constitutional problems. ACLU agrees that protecting “election integrity is a weighty governmental interest,” but under the First Amendment, “that interest may be accomplished . . . only by means that are narrowly tailored.” ACLU points to ample First Amendment case law holding that discussion “of public issues and debate on the qualifications of candidates are integral to the operation of the system of government established by our Constitution.” Quoting *Buckley v. Valeo* (1976), ACLU writes that the First Amendment affords “the broadest protection to such political expression in order to assure the unfettered interchange of ideas for the bringing about of political and social changes desired by the people,” and, quoting *New York Times v. Sullivan* (1964), ACLU notes that debate on public issues must remain “uninhibited, robust, and wide-open.” Moreover, ACLU notes, the First Amendment affords protection “to even allegedly false statements about public officials and public figures.” ACLU fears that faced with “the prospect of vetting millions of different posts to determine if they are ‘materially deceptive and digitally modified or created,’ many platforms may instead choose to aggressively censor or prohibit speech out of caution, including speech by candidates or relating to entire political topics.” Citing *Brown v. Entertainment Merchants Association* (2011) and *Burstyn v. Wilson* (1952), ACLU concludes that however much the technology may change, “the basic principles of freedom of speech and the press” do not vary with each new medium of communication. ACLU believes that the provisions of AB 2655, as currently drafted, “threaten to intrude on those rights and deter that vital speech.”

REGISTERED SUPPORT / OPPOSITION:

Support

California Initiative for Technology & Democracy (sponsor)
AFSCME California
Asian Americans Advancing Justice - Asian Law Caucus
Asian Americans and Pacific Islanders for Civic Empowerment
Asian Law Alliance
Bay Rising
California Clean Money Campaign
California Initiative for Technology & Democracy, a Project of California Common CAUSE
California State Sheriffs' Association
California Voter Foundation
Chinese Progressive Association
Courage California
Disability Rights California
Hmong Innovating Politics
Indivisible CA Statestrong
Inland Empire United
League of Women Voters of California

Partnership for the Advancement of New Americans
SEIU California
The Partnership for the Advancement of New Americans
Verified Voting

Opposition

ACLU California Action
Chamber of Progress
Computer and Communications Industry Association
Electronic Frontier Foundation
Internet.Works
NetChoice
Software & Information Industry Association
TechNet

Opposition unless amended

Oakland Privacy

Analysis Prepared by: Tom Clark / JUD. / (916) 319-2334

Exhibit 5

[Hearings](#)[Share](#)

Assembly Standing Committee on Judiciary

April 23, 2024

[Return to comment starting point](#)

Search transcript

Khara Boender
Person

Thank you, Madam Vice Chair, Members of the Committee. My name is Khara Boender, testifying on behalf of the Computer and Communications Industry Association in respectful opposition to AB 2655. CCIA is an international, not for profit trade association with about two dozen members from a range of communications and technology firms.

Khara Boender
Person

CCIA and its members take seriously the impact deceptive content may have on elections, and many of our members are working to implement tools to better detect and label AI generated content. Using a combination of AI and human review, they moderate content in violation of their terms of service, including content that is illegal and potentially harmful. But the tools that are currently available are not always reliable or accurate.

Khara Boender
Person

Just as spam filters sometimes mislabel legitimate emails. And while such technology is evolving, so are the means for bad actors to evade such detection. Because covered platforms are not privy to the intent and context for which a piece of content is used, they could inadvertently over block or over label content. This could result in user frustration and suppression of political speech. Political speech was at the core of why our First Amendment was established, and it is critical to maintain those protections.

Khara Boender
Person

Responsibility for labeling AI generated election content and liability for the deceptive content should rest with the entity that puts forth such material, the one that is most aware of the intent and context for which the content was created and shared. We also have concerns about the breadth of the bill. While it establishes defined time periods surrounding elections and election processes for newly established prohibitions and requirements, it

View Agenda

CURRENTLY DISCUSSING

Bill AB 2655

Defending Democracy from Deepfake Deception Act of 2024.

[View Bill Detail](#)[Download Transcript](#)

COMMITTEE ACTION: PASSED ✓

FOR 9 AGAINST 0 NO VOTE 3

[Brian Maienschein](#)[Tina McKinnor](#)[Ash Kalra](#)[Rebecca Bauer-Kahan](#)[Show all 9 votes](#)Next bill discussion: [May 22, 2024](#)Previous bill discussion: [April 10, 2024](#)

Speakers

LEGISLATOR

[Rebecca Bauer-Kahan](#)[Marc Berman](#)[Isaac Bryan](#)[Damon Connolly](#)

ADVOCATE

[Khara Boender](#)[Chanel Freeman](#)[David Harris](#)[Eric Harris](#)[Show more speakers](#)

Journalists backed by artificial intelligence bringing transparency and accountability to California's policy choices.

[About Digital Democracy](#)[Data & Methodology](#)[Visit CalMatters](#)

Support this nonprofit initiative

Digital Democracy informs Californians, holds officials accountable, and builds a stronger community. Brought to you by the 501(c)(3) nonprofit and nonpartisan CalMatters newsroom.

[Support Our Work](#)

Sign up for news and updates

Receive updates about Digital Democracy and CalMatters' daily newsletter that brings transparency to state government.

Email

[Sign Up](#)

By signing up, you agree to the [terms](#).

Assembly Standing Committee on Judiciary
AB 2655

Unidentified Speaker: Thank you.

Ash Kalra: Up next, we have item 12. AB 2655, Berman. We have a motion and a second. Before Assembly Member Berman even took a seat. He has a motion in a second.

Marc Berman: Thank you, colleagues. I'd like to start by thanking Committee staff for identifying areas needing work, including greater clarity between when something is blocked versus labeled, better defining election process and raising the bar on citizen suits. I look forward to working with the Committee to address these issues. I'm also happy to take the amendments on pages 9 and 10. And I thank the staff on the Privacy and Consumer Protection Committee for suggesting definitional amendments to align this Bill with others in the AI space.

Marc Berman: Five years ago, I authored the first election deepfake bill in the nation. To be honest, by the time that was signed into law, it was already too weak, and it's only gotten weaker since then. Just a few short years after that Bill passed, the technology is better, cheaper, and widely accessible. As a result, we're seeing deepfakes used to undermine elections across the globe and here in America and California.

Marc Berman: As technology changes, so too must our laws, and there's too much at stake to rely on nonbinding and voluntary agreements to police election deepfakes. Therefore, I'm authoring AB 2655 to protect election integrity by regulating the online spread of disinformation and deepfakes meant to influence an election with today's sophisticated deepfakes, voters may not know what images, audio, or video they can trust. This deceptive and hyper-realistic content can undermine faith in election integrity.

Marc Berman: Accordingly, AB 2655 would, for a limited period of time before and after an election, require large online platforms if they know or should know that the content meets the test in the Bill. To restrict the distribution of materially deceptive and digitally altered or created images, audio, or video meant to influence an election. For less harmful, yet still materially deceptive and modified content, the Bill would require the platforms to label other AI-created or modified election disinformation meant to influence.

Marc Berman: In both instances, the Bill does not demand perfection. What it says is that platforms must act if they know, or should know, using best available tools, that the content meets the test in the Bill. Platforms can't bury their head in the sand, but if they don't know, then there's no obligation. The Bill protects against false speech that has been intentionally manipulated to deceive and done with malice. This is defamation and an exception to the First Amendment.

Marc Berman: Additionally, it is important to note that in order to comply with Section 230 of the Communications Decency Act, the Bill does not provide for any monetary penalty against online platforms. Instead, the only penalty is that a court will order the platforms to do what the Bill requires. I realize that this Bill is

attempting to legislate in an arena where technology and the law are fast evolving, and that can be challenging on many levels.

Marc Berman: I can't promise that the Bill will survive a First Amendment or Section 230 court challenge. I do commit to continuing to make amendments to improve the Bill, to try to guard against that. But the Bill is crafted to give us the best chance to survive a strict scrutiny review. But the reality is that doing nothing isn't an option. Therefore, I respectfully request an aye vote.

Ash Kalra: Thank you.

Leora Gershenzon: Chair and Members. I'm Leora Gershenzon with the California Initiative for Technology and Democracy, a project of California Common Cause. As elected officials, you are all keenly aware of the threats posed by AI-generated election deepfakes. Today's Washington Post there's a headline that reads, AI deepfakes threaten to upend global elections. No one can stop them. We can't just be in that position.

Leora Gershenzon: AB 2655 strikes the right balance by banning, for a strictly limited time around elections, the online spread of the worst deepfakes intended to deceive voters and influence elections. Videos, images, and audio showing candidates, election officials, or voting apparatus doing or saying something they did not do or say. The Bill also requires other fake content regarding elections to be labeled as such, again, for a limited period of time around the elections.

Leora Gershenzon: The Bill only applies to the largest online platforms with the greatest reach of potential election disinformation and the deepest pockets to help find, to help determine what is and isn't AI-generated fake content. Assembly Member Berman has worked diligently to ensure that AB 2655's approach is narrowly tailored and does not extend to opinions or controversial statements. It only seeks at peak elections time to stop the use of demonstrably false content hurled at political leaders in an effort to disrupt our fair elections.

Leora Gershenzon: The Bill is respectful of the First Amendment by serving a compelling government interest, protecting our free and fair elections, and by being narrowly tailored to address that compelling interest in terms of time, subject matter, and requiring malice. It also limits remedies because there are no financial penalties because of Section 230. I would like to end very quickly with a quote from that article in the Washington Post from Senator Klobuchar, which said, let's not stand on the sidelines with our elections while our elections get screwed up.

Leora Gershenzon: This is like a hair-on-fire moment. This is not a let's wait and see. Let's wait three years and see how it goes moment.

Ash Kalra: Thank you.

David Harris: Chair and Members. My name is David Harris and I'm a senior policy advisor to PSYT IT. I'm also a chancellor's public scholar at UC Berkeley, where I've been teaching since 2015. I teach classes including Civic Technology and AI Ethics for Leaders. My students and I go over all of the ways that technology can

play a role in our democracy. I also previously worked for close to five years on civic integrity, social impact, and responsible AI at Facebook and Meta.

David Harris: I have advised the European Union, the White House and NATO on AI and the role that it plays in democracy. This Bill only applies to the largest online platforms with the greatest reach for potential election misinformation, and it is fully implementable today. Based on tools that these companies already possess for dealing with illegal content, such as child sexual abuse material or terrorist content, they could apply the same systems, techniques, procedures, and processes that they use for that type of content to managing deepfakes in elections.

David Harris: This Bill is narrow and in a narrow way makes certain types of content impermissible. Now, some of these companies will suggest that these requirements can't be met, but they are already generally subject to similar requirements under the European Union's Digital Services Act and the recently released guidelines that the EU has released about elections and how that, the Digital Services Act, applies to elections.

David Harris: Furthermore, 20 of the biggest tech companies involved committed at the Munich Security Conference in February to something called the AI elections accord, where they already promised to develop and implement technology to mitigate risks related to deceptive AI election content.

David Harris: The problem is that none of their commitments made in Munich are binding or have any timelines associated with them, and they cannot be held accountable. That's why this Bill, AB 2655 is what we need to assure that the tech companies live up to the promises that they have already made.

Ash Kalra: Thank you. Is there anyone else here in support of AB 2655?

Chanel Freeman: Chanel Freeman, on behalf of the League of Women Voters of California in strong support.

Bryant Miramontes: Bryant Miramontes with AFSCME California in support.

Eric Harris: Eric Harris, Disability Rights California, in support.

Diane Dixon: Do we have any speakers in opposition? Please come forward.

Khara Boender: Thank you, Madam Vice Chair, Members of the Committee. My name is Khara Boender, testifying on behalf of the Computer and Communications Industry Association in respectful opposition to AB 2655. CCIA is an international, not for profit trade association with about two dozen members from a range of communications and technology firms.

Khara Boender: CCIA and its members take seriously the impact deceptive content may have on elections, and many of our members are working to implement tools to better detect and label AI generated content. Using a combination of AI and human review, they moderate content in violation of their terms of service, including content that is illegal and potentially harmful. But the tools that are currently available are

not always reliable or accurate.

Khara Boender: Just as spam filters sometimes mislabel legitimate emails. And while such technology is evolving, so are the means for bad actors to evade such detection. Because covered platforms are not privy to the intent and context for which a piece of content is used, they could inadvertently over block or over label content. This could result in user frustration and suppression of political speech. Political speech was at the core of why our First Amendment was established, and it is critical to maintain those protections.

Khara Boender: Responsibility for labeling AI generated election content and liability for the deceptive content should rest with the entity that puts forth such material, the one that is most aware of the intent and context for which the content was created and shared. We also have concerns about the breadth of the bill. While it establishes defined time periods surrounding elections and election processes for newly established prohibitions and requirements, it is unclear when an election's process starts. Further, the bill does not specify which elections or where.

Khara Boender: And while the bill exempts satire and parody, it is unclear who gets to decide what constitutes those uses. Faced with individual users seeking injunctive relief merely if they disagree with a covered platform's decision regarding reported content, a service may choose to prohibit all digitally altered content, cutting off many valuable and helpful uses. These tools can be used by campaigns to reach voters with high quality content at lower costs or to translate speech into multiple languages. I'll wrap up there. Appreciate your time today in consideration of our comments, and welcome any questions.

Diane Dixon: Thank you. Go ahead.

Jose Torres Casillas : Thank you. Good morning, Madam Chair, and Mr. Chair, as he walks in. Jose Torres with TechNet. We are respectfully opposed to AB 2655. I echo the concerns of my colleague from CCIA and want to emphasize this bill requires online platforms to make determinations about truth and falsity in an impossible way. Instances where content and or information is clearly true or clearly false are not norm.

Jose Torres Casillas : Far more often, content falls into a middle ground where it requires time and effect intensive investigation to determine whether to determine whether something is true or false. Investigative journalists have challenges with fact checking even the most high profile races or candidates. It is difficult enough for a platform to know whether something is false as it relates to a presidential candidate or a high profile federal race, and this is simply impossible for races lower down on the ticket.

Jose Torres Casillas : A platform cannot accurately adjudicate reports on those types of content and will instead resort to over removing information in order to avoid liability and the penalties in this bill. Removing information that is only suspected of being false is clearly not a good outcome. And then for these reasons, we are respectfully opposed. Thank you.

Ash Kalra: Thank you. Is there anyone else here in opposition to AB 2655?

Cynthia Valencia: Cynthia Valencia with the ACLU California Action in opposition due to First Amendment concerns.

Ash Kalra: Thank you.

Tracy Rosenberg: Tracy Rosenberg on behalf of Oakland Privacy. We still have a couple of concerns. We thank the Committee for the amendments. We'll be checking them out. Thank you.

Ash Kalra: Thank you. All right. We'll bring it back to the Committee. Any questions? Assembly Member Bauer-Kahan.

Rebecca Bauer-Kahan: Hi. I know. So I have to start by saying thank you for your leadership in the elections arena. As we heard earlier about a different subject matter, there's no one I trust more to shepherd a bill of this magnitude through the Legislature. But I do have some questions because I agree, I think we all agree that strict scrutiny would be applied.

Rebecca Bauer-Kahan: I don't think, I know I don't have any question that there is a strong governmental interest, so that we don't need to discuss. So the question then to me is is this narrowly tailored in a sufficient way as to both pass scrutiny and to protect the First Amendment? I am a believer the First Amendment has immense value.

Rebecca Bauer-Kahan: And so one of the things that I actually think was valuable in the analysis was citing two year prior law, and so we could compare some of the differences in the way this was drafted. And there were some changes in the way this was drafted that raised concerns for me. And I'll say the first one is that any California resident can bring this case. I do think when we as candidates challenge speech, there is a, we are held to account by the voters. Right.

Rebecca Bauer-Kahan: We are responsible to them, and they can take it out on us at the ballot box. Whereas if we have sort of a third party out there challenging this, I think it is concerning to me that you would have a lot more spurious challenges and ones potentially that are not intended to go after what is truly deceptive speech. So I was curious about that. Do you want me to do all my questions? There's quite a few. Okay. Then the question is also on the standard, right.

Rebecca Bauer-Kahan: The analysis talks about having a clear and convincing standard versus just the likelihood of success on the merits for the injunctive relief. I do think the higher standard given, again, that we need to narrowly tailor this to ensure that we are not looking at some case that might win, but instead something where there is clear and convincing evidence that it falls under the statute makes sense. So I don't know if you want to address that.

Rebecca Bauer-Kahan: But then as I looked at the question of what was deceptive material, what would fall under this, I found the definition of that incredibly broad. In the prior statute that you authored, it used a reasonable person standard

to decide whether something was potentially viewed as inauthentic. So the reasonable person that would view it, I have it right here. It said that would have falsely appeared to be a reasonable person, to be inauthentic, would falsely appear to be authentic to a reasonable person.

Rebecca Bauer-Kahan: And here we have, you know, the authenticity question even has, or that otherwise generates an inauthentic image. That is anything that is created using Gen AI. So now anything that is created using Gen AI would have an inauthentic image and would fall under this. And then it says that, you know, that contains a false portrayal of any of the following in candidate. You know, what they said or did, which is what you said.

Rebecca Bauer-Kahan: And I was thinking about that and I thought, well, you know, we have a colleague and some allege that she had tried to allow for infanticide in California. I would argue that's not what she did. She would argue that's not what she did. But it's a fair argument in the public sphere. The First Amendment would protect whether what she did was or was not. I think that is protected First Amendment speech. But arguably somebody could go and say, that's not what I did. And we would be...

Ash Kalra: Oh, no

Marc Berman: It's all right. We're good.

Rebecca Bauer-Kahan: And we would be in a space where we have...

Marc Berman: I have faith.

Ash Kalra: Last call.

Rebecca Bauer-Kahan: I thought it was the, I thought it was the AI shutting us down. David can tell us if that's possible. So I guess my concern is that that is so broadly drafted that ads or communications that are meant to attack things that we arguably did but maybe we think we didn't do, would then be under subject to this 36 hours and taken down. And I just think that there are much stricter standards we could put on this to make sure we are talking about what I think you're actually trying to get at.

Rebecca Bauer-Kahan: Right, which is where people are actually being deceived to believe something, someone said something, a speech that looks authentic and a reasonable person would think that Joe Biden gave that speech, not a communication that is arguably an ad saying we did something using an inauthentic image because AI generated it, which I think is much past the point of what you are trying to regulate here. And I think the definitions don't just capture the stuff we're talking about, which is misinformation, disinformation, but goes beyond that to political speech that I think I would hope all of us would want to protect. So I guess if you want to address that.

Marc Berman: Yeah, no, thank you very much for the points. Thank you for the conversations that we've had in the past 18 hours. Let me first say for the first

two issues that you raised, we are actively investigating what makes sense and actively discussing with stakeholders, with opponents, with the sponsors, with interested legislators, and actively looking at whether or not we should narrow who can bring suit or who can bring a complaint, actively looking at a lot of other components of the bill.

Marc Berman: So needless to say, everything is still very much up for discussion and would love to continue those discussions with you. I'd love to continue those discussions with the opposition to try to make sure that, at the end of the day, if this bill is signed into law, that it is as narrowly tailored as possible. Because like you said, that's critically important to the bill, you know, surviving in the court. I'm going to defer to the experts on some of the specific language and what we still have and what we've changed. But, you know, this is very much a work in progress.

Leora Gershenzon: Yes, we did. We absolutely did not take out the false portrayal. So there's language on page five that says this is under the definition of digitally manipulated speech. False portrayal means the content would cause a reasonable person to have a fundamentally different impression or understanding of the content than if the person would have seen the authentic image. So it still keeps that same provision of false portrayal that was in the original Berman bill.

Rebecca Bauer-Kahan: But isn't that an or? Isn't that one of the options?

Leora Gershenzon: No, no, no. It's a requirement. A. You have to have under two. You have to have that. So false portrayal is a key part of this. So, in fact, we had discussions about taking that out and did not because that's clear. It has to be deceptive. So it has to be false. It has to be digitally manipulated, and it has to be false to, it has to deceive a person. And it isn't just the same like, you, infanticide. You cause this. But if there was a picture of the Member actually murdering a child, that would be, that would fall under this bill. Not just saying this author did a bill that included.

Rebecca Bauer-Kahan: Right. But the reasonable person only has to have a fundamentally different understanding or impression of the content or... It doesn't feel like the reasonable person has to believe that it is authentic. That is not how I read this. Is that how you read it?

Leora Gershenzon: Yes, because it has to appear authentic. We can tighten up the language. I mean, as the Assembly Member said, is this perfect language? It isn't. There's a lot more process to go through and there are definitely ways we can tighten this. But there are key things, which is that it is digitally manipulated. It is false. It is intended to deceive, the person who the original content producer intended to deceive. It does deceive people. And that it's intended falsely to influence the election.

Rebecca Bauer-Kahan: Right. So I see a reasonable person under false betrayal. My problem is you don't have a reasonable person under the authenticity question. That's my issue.

Leora Gershenzon: So you want...

Rebecca Bauer-Kahan: So I want. I think that the reasonable person has to believe this is authentic. Not just that it... Cause there you have an or it otherwise generates an inauthentic image. And so I think that I want someone to be viewing this to believe this is an authentic image. And I think that the way this is drafted with all of those requirements, the false portrayal, again, I think that a lot of the way people portray us, we would argue was false.

Rebecca Bauer-Kahan: I just, I mean, you know, that is what. And we have to accept that. We are public officials. That is part of the First Amendment protections. And so I also think that the image needs to be one such that, again, the viewer believes this is an authentic speech, it's an authentic image, et cetera.

Marc Berman: Totally appreciate the point. Absolutely.

Ash Kalra: Assembly Member Bryan.

Isaac Bryan: I'll be quick. Mr. Berman, when you brought this bill to me, I remember your pitch pretty well.

Marc Berman: I remember your response.

Isaac Bryan: It's supported by the Sheriff, it's opposed by the ACLU, and it may violate the First Amendment, but...

Marc Berman: You're paraphrasing.

Isaac Bryan: But similar to my colleague from the Bay Area, all of your past work in this space, when a bill analysis cites your own past work, it speaks to your dedication to the issue and your willingness and ability to navigate complex policy conversations like this one. And I think this is one of those conversations that needs to continue. So happy to give you a courtesy aye vote today, and looking forward to seeing where you ultimately take this work.

Marc Berman: Appreciate it.

Ash Kalra: All right, do we have a motion in a second? Assembly Member Connolly.

Damon Connolly: A couple quick questions. And yeah, I appreciate that sentiment as well. One question that was raised is why does it only target large platforms?

Marc Berman: I think we're trying to target the biggest risk, and the biggest risk, the biggest risk and the platforms that have the resources and the technological ability to be able to determine what content would apply. So I'll defer... At a high level, that's the reasoning. But defer to the experts to dive into the details a little bit.

Leora Gershenzon: In the ideal world, we would cover every single platform. But again, what we are asking of the platform to do does they've got to use state of the

art tools. The largest platforms have the capacity. The largest platforms also have the largest reach. So if you're trying to stop disinformation, we're never going to fully stop it. And by the way, I'm sorry I didn't mention, we're not trying to stop misinformation. That actually is much more allowable under the First Amendment.

Leora Gershenzon: It's the disinformation, it's the absolutely false context. So we're trying to go where the greatest tools are and not run into... It's impossible for these smaller startups. We are pro-innovation. We don't want to stop the tools that are needed to do this.

David Harris: If I could add, we chose to go for the larger platforms, actually, as a bit of a compromise. We looked at the way that the European Union defined very large online platforms and we actually went lower than that. They were looking at platforms that have more than 45 million, or 10% of the EU population approximately, as users. We went for a 1 million number within California, proportionately smaller. We didn't go lower than that because we do want to recognize that this is work. This has costs associated with it that are incidental to companies that are bringing in billions of dollars each quarter, but that are significant for small startups, and we wanted to acknowledge that.

Damon Connolly: And who decides whether something has been modified? Is it actually like an algorithm that does that, or is it people? What's that rigorous process?

Marc Berman: I'm going to defer to David, but algorithms, and then maybe backed up by people.

David Harris: Yeah, exactly. As is the case with almost all types of content moderation currently, there are two ways that a piece of content that can be enqueued for a moderation decision. One is in an automated fashion where an algorithm or AI system detects that something might contain child sexual abuse material, detects that it might contain a known false statement that has been previously fact checked, or detects that it is repeat of content that has been taken down before of any type.

David Harris: And so that can be done algorithmically. There are situations where an algorithm might also generate a unclear response and queue that automatically for a human reviewer. And then there are also ways that, if content passes through the algorithmic methods, it could still be reported by anyone, yourself included, through any platform's tools for reporting content. And then there's a possibility that you would have an algorithmic assessment or an AI assessment after the human reports. And that could, depending on how it unfolds, go back to a human actual reviewer. And there are in many of the platforms, multiple layers of human review that can be appealed.

Ash Kalra: I was just going to ask if we can start to wrap up. We have a lot of bills going through. So quickly, and quick, quick response, please.

Damon Connolly: Final quick question. So, and following up on that. The way the bill is currently set up is if an individual disagrees, quote unquote, with a platform response, a private lawsuit is possible. So I'm trying to get a sense of, that

sounds like kind of opening the floodgates potentially to litigation. But how is that envisioned? How is that going to play out realistically?

Leora Gershenzon: I mean, the Assembly Member has already discussed looking at ways to limit that, including who may bring the lawsuit and raising the standard from preponderance of the evidence to potentially clear and convincing. The penalty against the platforms. There's our litigation. These are lawsuits. So it's not that they're so easy, but there's no monetary penalties. So because of Section 230, they pay nothing. If they lose, the court will order them to simply do something that they hadn't done. So they will. They are simply ordered to comply with the law. There are no monetary penalties.

Damon Connolly: Well, I'm glad to hear though that there's further discussion about how to refine that.

Marc Berman: There is, absolutely.

Ash Kalra: Thank you. There is a motion already, and so would you like to close Assembly Member Berman?

Marc Berman: Respectful ask for an aye vote. I appreciate the conversation.

Ash Kalra: Thank you.

Committee Secretary: Do passes amended to Appropriations. [Roll Call]

Ash Kalra: All right, that will be placed on call.

Marc Berman: Thank you.

Ash Kalra: All right, so it's another check.

Committee Secretary: We have a vote change. Dixon, no to not voting.

Ash Kalra: Okay. All right. Assembly Member Bauer-Kahan, we have less than an hour to do six bills. If we can. If we can't, we're gonna have to come back either tonight or at some other moment sometime to finish these bills this week.

Rebecca Bauer-Kahan: I won't ask any questions this time, so it'll go faster. Thank you, Mr. Chair and Members, I'm proud to present AB 2930. Oh, and I'll be accepting the Committee amendments. Proud to present AB 2930. I want to thank the Committee for their very thorough analysis. It was incredibly well done. I appreciate it. As noted, it was modeled after President Biden's AI Bill of Rights. And AB 2930 is incredibly, incredibly simple, although the opposition would disagree.

Exhibit 6

Date of Hearing: April 10, 2024

ASSEMBLY COMMITTEE ON ELECTIONS
Gail Pellerin, Chair
AB 2655 (Berman) – As Amended April 1, 2024

SUBJECT: Elections: deceptive audio or visual media.

SUMMARY: Requires large online platforms, as defined, to block the posting or sending of materially deceptive and digitally modified or created content related to elections, or to label that content, during specified periods before and after an election. Specifically, **this bill**:

- 1) Requires a large online platform (platform), using state-of-the-art, best available tools to detect digitally modified or created content, to develop procedures for blocking and preventing the posting or sending of materially deceptive and digitally modified or created content, and to block and prevent that content if the platform knows or should know that the content meets the following requirements, during a specified time period, of any of the following:
 - a) A candidate portrayed as doing or saying something that the candidate did not do or say.
 - b) An elections official portrayed as doing or saying something in connection with the performance of their elections-related duties that the official did not do or say.
 - c) An elected official portrayed as doing or saying something that influences the election that the elected official did not do or say.
 - d) A voting machine, ballot, voting site, or other property or equipment related to an election that is portrayed in a materially false way.
- 2) Prohibits a platform from preventing a candidate, notwithstanding the prohibition detailed above, from portraying themselves as doing or saying something that the candidate did not do or say if the digital content includes a disclaimer stating: “This (image/video/audio) has been manipulated.” Requires this disclaimer to comply with the following:
 - a) In the case of visual media, requires the text of the disclaimer to appear in a size that is easily readable, as specified.
 - b) In the case of a video, requires the disclaimer to appear for the duration of the video.
 - c) In the case of media that consists of audio only, requires the disclaimer to be read in a manner that can be easily heard by the average listener at both the beginning and the end of the audio. For audio that is longer than two minutes, the disclaimer must also be included during the audio at intervals of not more than two minutes each.
- 3) Provides that the provisions detailed above apply only during the following time periods:
 - a) Beginning 120 days before any election through the day of the election.

- b) With respect to content pertaining to elections officials, or that depicts or pertains to elections equipment and materials, beginning 120 days before the election and ending on the 60th day after the election.
- 4) Requires a platform, using state-of-the-art, best available tools to detect digitally modified or created content, to develop procedures for labeling materially deceptive and digitally modified or created content that pertains to election processes and that is not subject to the blocking provisions outlined above, and to label such content as inauthentic, fake, or false if the platform knows or should know that the digitally modified or created content meets the requirements of this bill.
- a) Requires the label to permit users to click or tap on it and to inspect all available provenance data about the digitally modified or created content in an easy-to-understand format.
 - b) Provides that the labeling requirement detailed above applies only during the following time periods:
 - i) The period beginning one year before the election and through the day of the election that is specified in or implicated by the content.
 - ii) The period beginning one year before the election process and through the final day of the election process that is specified in or implicated by the content.
 - iii) If the content depicts or pertains to elections officials, the period beginning one year before the election or election process that is specified in or implicated by the content and ending on the 60th day after that election or the 60th day after the final day of that election process, as applicable.
- 5) Requires a platform to provide a way for Californians to report content that was not blocked or labeled as required. Requires the platform to respond within 36 hours and to describe any action taken.
- 6) Permits a Californian who reported content and who does not receive a response in 36 hours or who disagrees with the response, the Attorney General (AG), a district attorney, or a city attorney to seek injunctive or other equitable relief against a platform to compel compliance with this bill. Requires the court to award a prevailing plaintiff reasonable attorney's fees and costs. Provides that such an action is entitled to precedence in court, as specified.
- 7) Requires any label or disclaimer that must appear on content under this bill to appear in English, and in the same language as the content if the content isn't in English.
- 8) Requires a platform to maintain a copy of any content that it blocks or labels under this bill for at least five years from the election or election process specified or implicated in the content. Requires the platform to make that content available to the Secretary of State, the Fair Political Practices Commission, and researchers, if requested.
- 9) Provides that this bill does not apply to any of the following:

- a) A regularly published online newspaper, magazine, or other periodical of general circulation that routinely carries news and commentary of general interest, and that publishes any materially deceptive and digitally altered or digitally created image, audio, or video recording that an online platform is required to block or label by this bill, if the publication contains a clear disclosure that the materially deceptive and digitally altered or digitally created image or audio or video recording does not accurately represent any actual event, occurrence, appearance, speech, or expressive conduct.
- b) Materially deceptive audio or visual media that constitutes satire or parody.

10) Defines the following terms, for the purposes of this bill:

- a) “Election processes” to mean any government process related to an election, including, but not limited to, elections, candidates, vote counting, redistricting, and proceedings or processes of the electoral college.
- b) “Materially deceptive and digitally modified or created content” to mean an image or an audio or video recording or other digital content, including a chatbot, that has been intentionally manipulated such that all of the following conditions are met:
 - i) The digital content is the product of digital manipulation, artificial intelligence (AI), or machine learning, including deep learning techniques, that merges, combines, replaces, or superimposes content onto an image or an audio or video recording, creating an image or an audio or video recording that appears authentic, or that otherwise generates an inauthentic image or an audio or video recording that appears authentic.
 - ii) The content contains a false portrayal of a candidate for elective office, an elected official, an elections official, or a voting machine, ballot, voting site, other property or equipment related to an election, or elections process. Provides that a “false portrayal” means the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if they were hearing or seeing the unaltered, original version of the content.
 - iii) The person or entity who attempted to post or send, or who did post or send, the content did so knowing the portrayal was false, or did so with reckless disregard for whether the portrayal was false. Provides that if the content is intentionally manipulated and contains a false portrayal, there is a rebuttable presumption that the person or entity knew the portrayal was false or that they acted with reckless disregard for whether the portrayal was false.
- c) “Large online platform” to mean a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1 million California users during the preceding 12 months.

- 11) Provides that content that contains only minor modifications that do not lead to significant changes to the perceived contents or meaning of the content are not “materially deceptive and digitally modified or created content” for the purposes of this bill, as specified.
- 12) Contains various findings and declarations and contains a severability clause.

EXISTING STATE LAW:

- 1) Prohibits a person, committee, or other entity, until January 1, 2027, from distributing with actual malice, within 60 days of an election at which a candidate for elective office will appear on the ballot, materially deceptive audio or visual media of a candidate with the intent to injure the candidate’s reputation or to deceive a voter into voting for or against the candidate.
 - a) Defines “materially deceptive audio or visual media,” for these purposes, as an image or an audio or visual recording of a candidate’s appearance, speech or conduct that has been intentionally manipulated in a manner that both of the following are true about the image or audio or video recording:
 - i) It would falsely appear to a reasonable person to be authentic; and,
 - ii) It would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than the person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording.
 - b) Provides that this prohibition does not apply if the audio or visual media includes a disclaimer stating “This (image/video/audio) has been manipulated,” and the disclaimer complies with specified requirements.
 - c) Permits a candidate whose voice or likeness appears in deceptive audio or visual media distributed in violation of this provision to seek the following relief:
 - i) Injunctive or other equitable relief prohibiting the distribution of the materially deceptive audio or visual media in violation of this bill. Provides that such an action is entitled to precedence in court, as specified.
 - ii) General or special damages against the person, committee, or other entity that distributed that audio or visual media. Permits the court to award reasonable attorney’s fees and costs to a prevailing party in such an action.
 - d) Provides that in any civil action brought pursuant to these provisions, the plaintiff bears the burden of establishing the violation through clear and convincing evidence.
 - e) Provides that this prohibition shall not be construed to alter or negate any rights, obligations, or immunities of an interactive service provider under Section 230 of the federal Communications Decency Act.
 - f) Provides that this prohibition does not apply to any of the following:

- i) A radio or television broadcasting station, as specified, in either of the following circumstances:
 - (1) When it broadcasts materially deceptive audio or visual media as part of a bona fide newscast, news interview, news documentary, or on-the-spot coverage of bona fide news events, if the broadcast clearly acknowledges through content or disclosure that there are questions about the authenticity of the audio or visual media, as specified.
 - (2) When it is paid to broadcast materially deceptive audio or visual media.
 - ii) An internet website, or a regularly published newspaper, magazine, or other periodical of general circulation, including an internet or electronic publication, that routinely carries news and commentary of general interest, and that publishes materially deceptive audio or visual media covered by this prohibition, if the publication clearly states that the media does not accurately represent the speech or conduct of the candidate.
 - iii) Materially deceptive audio or visual media that constitute satire or parody. (Elections Code §20010, as amended by Section 3 of Chapter 745 of the Statutes of 2022)
- 2) Prohibits a person, firm, association, corporation, campaign committee, or organization, beginning January 1, 2027, with actual malice, from producing, distributing, publishing, or broadcasting campaign material, as defined, that contains either of the following types of pictures or photographs, as specified, unless the campaign material includes a disclaimer that the picture is not an accurate representation of fact:
- a) A picture or photograph of a person or persons into which the image of a candidate for public office is superimposed.
 - b) A picture or photograph of a candidate for public office into which the image of another person or persons is superimposed. (Elections Code §20010, as amended by Section 4 of Chapter 745 of the Statutes of 2022)

EXISTING FEDERAL LAW provides, pursuant to Section 230 of the federal Communications Decency Act, that no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. §230)

FISCAL EFFECT: Unknown

COMMENTS:

- 1) **Purpose of the Bill:** According to the author:

AB 2655 will ensure that online platforms restrict the spread of election-related deceptive deepfakes meant to prevent voters from voting or to deceive them based on fraudulent content. Deepfakes are a powerful and dangerous tool in the arsenal of those that want to wage disinformation campaigns, and they have the potential

to wreak havoc on our democracy by attributing speech and conduct to a person that is false or that never happened. Advances in AI make it easy for practically anyone to generate this deceptive content, making it that much more important that we identify and restrict its spread before it has the chance to deceive voters and undermine our democracy.

- 2) **Threats of Manipulated Media in Campaign Communications:** The use of false and deceptive information in campaigns to influence election outcomes is not a new phenomenon. Laws aimed at curbing such practices and preserving the integrity of elections have a long history in California. In 1850, the First Session of the California State Legislature created penalties for election misconduct, including for “deceiving [an elector] and causing him to vote for a different person for any office than such elector desired or intended to vote for” (Chapter 38, Statutes of 1850). California law today includes various provisions criminalizing deceptive tactics that undermine election integrity or interfere with voters’ ability to participate in elections. This includes laws that prohibit distribution of false and misleading information about qualifications to vote or about the days, dates, times, and places where voting may occur (Elections Code §18302); prohibit the misleading use of government seals in campaign literature (Elections Code §18304); and prohibit coercing or deceiving people into voting in a way that was inconsistent with the person’s intent (Elections Code §§18573, 18573.5).

Advancements in technology have made it increasingly simple to produce false and misleading media that closely resembles authentic content. Moreover, platforms like social media have facilitated the rapid dissemination of deceptive media to large audiences at minimal cost. Given these developments, the potential threat posed by manipulated media to future elections’ integrity may be more significant than in the past.

As described in greater detail below, past legislative efforts have addressed concerns about manipulated media’s use to deceive voters during elections. Those laws, however, are limited, and are designed primarily to target the harms to *candidates* that may result from the distribution of manipulated media of those candidates. In contrast, this bill aims to regulate materially deceptive and digitally altered media depicting not only candidates, but also elections officials and elected officials who are not candidates. Additionally, this bill targets media that portrays elections materials and equipment in materially deceptive ways. The author and supporters of this bill believe that these provisions will safeguard voters against deceitful media that could undermine trust in the electoral process.

- 3) **Recent Examples of Materially Deceptive Campaign Communications:** As evidence of the need for this bill, the author points to the following incidents, as reported in the media:
- AI tools were used to create deepfake video ads of British Prime Minister Sunak on Facebook.
 - A Chinese disinformation campaign in Taiwanese elections made use of deepfakes, and AI-generated videos, images and audio clips.
 - AI tools additionally were used to disrupt elections in Argentina, Bangladesh, Pakistan, Slovakia.

- Governor Ron DeSantis created a deepfake of former President Trump hugging Anthony Fauci.
 - AI was used by a democratic operative and a magician to generate deepfake audio of President Biden's voice that was used in robocalls to dissuade voters from voting in the primary election.
- 4) **Previous Legislation Related to Materially Deceptive Media in Campaigns:** In 2019, in response to concerns that deepfake technology could be used to spread misinformation in political campaigns, the Legislature approved and Governor Newsom signed AB 730 (Berman), Chapter 493, Statutes of 2019. Deepfake technology refers to software capable of producing a realistic looking video of someone saying or doing something that they did not, in fact, say or do. This technology has advanced rapidly in recent years thanks to the use of AI to help train the software.

AB 730 prohibits the distribution of materially deceptive audio or visual media with actual malice with the intent to injure a candidate's reputation or to deceive a voter into voting for or against a candidate, unless the materially deceptive audio or visual media includes a disclaimer that it has been manipulated. AB 730 does not apply exclusively to deepfakes, but rather applies to any intentional manipulation of audio or visual images that results in a version that a reasonable observer would believe to be authentic. Nonetheless, the increasing availability and advancing capability of deepfake technology was the immediate impetus for that bill.

AB 730 was designed as an update to California's "Truth in Political Advertising Act," a law enacted in 1998 (through the passage of AB 1233 (Leach), Chapter 718, Statutes of 1998) that prohibited campaign material that contains a picture of a person into which a candidate's image is superimposed, or contains a picture of a candidate into which another person's image is superimposed, except if a specified disclaimer was included. The Truth in Political Advertising Act was introduced in response to the use of photoshopped pictures in campaign materials, and accordingly was designed to target the manipulation of photographs in campaign materials. In the 20 years following its passage, however, it was never amended to update the law to address more modern techniques of manipulating campaign materials in a manner that can mislead voters. AB 730 replaced the Truth in Political Advertising Act with a law that regulates not only altered photographs in campaign materials, but also audio and video media that have been altered in a materially deceptive manner.

AB 730 included a January 1, 2023 sunset date. In 2022, however, the Legislature approved AB 972 (Berman), Chapter 745, Statutes of 2022, which extended the sunset date to January 1, 2027. AB 972 did not otherwise change the provisions of AB 730. If the current January 1, 2027 sunset date is not repealed or extended, the original Truth in Political Advertising Act as enacted by AB 1233 of 1998 would go back into effect.

Because the impetus for AB 730 was concern about the potential that people might create deepfake media appearing to be accurate representations of the conduct of candidates for office, its provisions apply exclusively to images or audio or video recordings of a candidate's appearance, speech, or conduct. Relatedly, candidates for elective office who are the target of materially deceptive media are the only entities that can seek injunctive relief or damages under AB 730. Materially deceptive images, audio, or video that appear in

campaign communications are not covered by AB 730 if that media is not of a candidate. For instance, if a candidate digitally manipulated video or a photo of a campaign rally to make the crowd look significantly larger than it actually was, such manipulation would not be covered by AB 730 as long as the manipulated image or video was not materially deceptive about a candidate's appearance, speech, or conduct. Similarly, manipulated and materially deceptive content in advertisements related to ballot measures, or in communications that seek to undermine confidence in the electoral process but that do not mention candidates directly, generally would not be covered by AB 730.

- 5) **Online Platforms and Materially Deceptive Content:** This is one of two bills being considered by the committee today that seek to address materially deceptive and digitally altered elections-related content in an effort to protect the integrity of elections in California. While the related bill that is being considered today (AB 2839 (Pellerin)) applies broadly to the distribution of such content through various mediums, this bill specifically targets the distribution of deceptive content through online platforms, including social media. Recognizing that those online platforms can facilitate the rapid spread of deceptive content, this bill seeks to minimize that potential by obligating large online platforms to block or label offending content. In order to do that, the platforms necessarily will need to be able to identify the content that must be blocked or labeled.

In recognition that the regulation of the distribution of content can create free speech concerns, this bill contains various provisions that tailor the content to which it applies, such that it targets content that has the highest likelihood of deceiving voters and undermining electoral integrity. While that tailoring does limit the content that online platforms would be required to block or label, it also adds additional factors that platforms must consider in order to identify content that is required to be blocked or labeled under this bill.

Along with other limitations, this bill applies only to content that (1) is distributed during specified time periods around elections and election processes, (2) includes media relating to elections or the electoral process in specified ways, (3) that was intentionally manipulated digitally to be materially deceptive, and (4) that is not satire or parody. Each of these limitations adds additional factors that online platforms would need to consider when determining whether a specific communication must be blocked or labeled by this bill.

For example, in order to determine whether it must block content that *portrays a candidate for election as doing or saying something that the candidate did not do or say*, the platform would need to know not only that the person portrayed in the content was a candidate for office, but also the date (or dates) of the election when the candidate will appear on the ballot. Similarly, it would need to determine whether the candidate had actually said or done the thing that the candidate is portrayed as doing. While some of that information will be widely available and well known in some cases (e.g., the identity of major party candidates for President of the United States in presidential general elections and the dates of federal elections), it will be more arcane in other situations. Given the number of elections (including standalone local and special elections) and candidates (including write-in candidates and candidates for local elections in smaller jurisdictions) in California at any given time, making the determinations at scale about which content must be blocked or labeled likely will be considerably more challenging than making those determinations on a case-by-case basis in a court of law.

This bill includes numerous provisions that recognize that platforms will face challenges in making some of these determinations, and in limiting those platforms' obligations and liabilities accordingly. For instance, recent amendments to this bill provide for the large online platforms to use "state-of-the-art, best available tools" for detecting digitally modified or created content, recognizing that the identification of such content with perfect accuracy is impossible. Other recent amendments provide that an online platform is obligated to block or label content only if the platform "knows or should know" that the content meets the requirements of the bill, and provide that the bill's requirements relating to content portraying an elections official applies only if the platform knows or should know that the person is an elections official. Furthermore, unlike the related legislation that targets materially deceptive election content more broadly, this bill does not provide for damages to be awarded against platforms that fail to comply with their obligations. Instead, the only legal relief available under this bill is injunctive relief, and a person (other than the AG, a district attorney, or a city attorney) would be required to report the content to the platform and give the platform an opportunity to block or label the content before the person could seek that relief. Notwithstanding these provisions, the extent to which large online platforms will be able to accurately block and label materially deceptive elections-related content at scale, as contemplated by this bill, is unclear.

- 6) **Free Speech Considerations:** The First Amendment to the United States (US) Constitution, which also applies to states under the Fourteenth Amendment, provides in relevant part "Congress shall make no law...abridging the freedom of speech..." Similarly, Section 2 of Article I of the California Constitution provides in relevant part "Every person may freely speak, write, and publish his or her sentiments on all subjects, being responsible for the abuse of this right. A law may not restrain or abridge liberty of speech or press."

This bill seeks to regulate the distribution by online platforms of media containing intentionally manipulated images, audio, or video related to candidates, elections officials, elected officials, and election materials and equipment under certain circumstances. A question could be raised about whether this bill is consistent with the right to freedom of speech that is guaranteed by the US and California constitutions. The US Supreme Court has ruled that even false statements are protected by the First Amendment (*United States v. Alvarez* (2012), 567 U.S. 709). When a law burdens core political speech, the restrictions on speech generally must be "narrowly tailored to serve an overriding state interest," *McIntyre v. Ohio Elections Commission* (1995), 514 US 334.

This bill targets deceptive content that could undermine trust in elections, prevent voters from voting, and distort the electoral process. The US Supreme Court generally has found that the protection of the integrity of elections is an overriding (or compelling) government interest (*Id.* at 349; *Burson v. Freeman* (1992) 504 U.S. 191, 199). A challenge of this bill on First Amendment grounds, then, likely would hinge on whether the court found this bill's provisions to be narrowly tailored.

As discussed in more detail above, this bill includes provisions to limit its scope to communications posing the greatest threat to election integrity. Whether these limitations adequately protect this bill from a potential constitutional challenge is a question that falls more squarely within the jurisdiction of the Assembly Judiciary Committee, where this bill will be heard next if it is approved by this committee. However, while these limitations may help protect the bill against a constitutional challenge, they may also make it harder for the

bill to achieve its aims of limiting the spread of materially deceptive communications that have the potential to undermine election integrity.

- 7) **Arguments in Support:** The sponsor of this bill, the California Initiative for Technology & Democracy, a Project of California Common Cause, writes in support:

Those trying to influence campaigns – conspiracy theorists, foreign states, online trolls, and candidates themselves – are already creating and distributing election-threatening deepfake images, audio, and video content in the US and around the world. This threat is not imaginary: generative AI has been used in various ways – most of them deeply deceptive – to influence the national elections in Slovakia, Bangladesh, Argentina, Pakistan, and elsewhere, including in our own country. Examples of this occurring in U.S. elections include Ron Desantis using AI-generated images to attack his opponent in his presidential run, foreign states caught attempting to influence American politics through social media, and just this month, a supporter of former President Trump creating a deepfake image of Trump with Black Americans designed to persuade Black voters to support Trump...

[AB] 2655 strikes the right balance by seeking to ban, for a strictly limited time before and after elections, the online spread of the worst deepfakes and disinformation maliciously intended to prevent voters from voting or getting them to vote erroneously based on fraudulent content. The bill also requires that other fake online content related to elections and elections processes (such as redistricting), which is also designed to undermine election procedures and democratic institutions, must be labeled as fake, again just for a limited time. The bill only applies to the largest online platforms with the greatest reach of potential election disinformation, and we believe it is fully implementable today based on tools these companies already possess. The companies covered by the bill's requirements are all already subject to similar requirements under the European Union's Digital Services Act, which is designed to, among other things, crack down on election interference.

AB 2655's approach is narrowly tailored and does not extend the law to hot button controversies or inflammatory claims – it does not ask social media platforms to adjudicate controversial opinions post by post. It simply stops the use of obviously, demonstrably untrue and provably false content meant to impermissibly influence our elections at peak election times.

- 8) **Arguments in Opposition:** A joint letter of opposition submitted by Chamber of Progress, Computer and Communications Industry Association, NetChoice, Software & Information Industry Association, and TechNet states:

Responsible digital services providers take aggressive steps to moderate dangerous and illegal content, consistent with their terms of service. The companies deliver on the commitments made to their user communities with a mix of automated tools and human review. In 2021, a number of online businesses announced that they had been voluntarily participating in the Digital Trust & Safety Partnership (DTSP) to develop and implement best practices to

ensure a safer and more trustworthy internet, and have recently reported on the efforts to implement these commitments.

AB 2655 appears to be based on the false assumption that online platforms definitively know whether any particular piece of content has been manipulated in such a way that is defined under the bill. While digital services may employ tools to identify and detect these materials with some degree of certainty, it is an evolving and imperfect science in its current form. AB 2655 also presumes that online platforms are an appropriate arbiter of deciding what constitutes accurate election information. However, most digital services are not equipped with the tools or expertise to make such judgments.

AB 2655 would impose significant operationally and practically challenging requirements on large online platforms who may not be best suited to achieve the bill's laudable and important goal of ensuring California's elections remain free and fair.

- 9) **Related Legislation:** AB 2355 (Wendy Carrillo), which is also being heard in this committee today, requires a political advertisement that is generated in whole or in part using AI to include a disclaimer stating that fact.

AB 2839 (Pellerin), which is also being heard in this committee today, prohibits the distribution of campaign advertisements and other election communications that are materially deceptive and digitally altered or created, except as specified.

- 10) **Double-Referral:** This bill has been double-referred to the Assembly Judiciary Committee.

REGISTERED SUPPORT / OPPOSITION:

Support

California Initiative for Technology & Democracy, a Project of California Common Cause
(Sponsor)

Asian Americans Advancing Justice - Asian Law Caucus

Asian Americans and Pacific Islanders for Civic Empowerment

Asian Law Alliance

California Clean Money Campaign

California State Sheriffs' Association

California Voter Foundation

Courage California

Disability Rights California

Indivisible CA Statestrong

Inland Empire United

The Partnership for the Advancement of New Americans

Verified Voting

Opposition

Chamber of Progress

Computer and Communications Industry Association

Electronic Frontier Foundation
NetChoice
Oakland Privacy (unless amended)
Software & Information Industry Association
TechNet

Analysis Prepared by: Ethan Jones / ELECTIONS / (916) 319-2094

Exhibit 7

Assembly Standing Committee on Elections
AB 2655

Gail Pellerin: Perfect, perfect. Your second Bill. So you're AB 2655.

Marc Berman: Great. Just give a second for my witnesses, assuming I have witnesses, and I do hello. So five years ago, as some folks who were in the Legislature at that time might remember, I authored the first election related deepfake Bill in the country. I did this after watching Jordan Peele, the actor and director, created a deepfake of President Obama in 2018.

Marc Berman: And it was funny, but it also really set off alarm bells for me as it was clear and easy to see how this promising technology could be abused by bad actors, especially to influence our elections. However, the potential harm was fairly limited back then as the technology to make a deepfake was very expensive. It wasn't widely available.

Marc Berman: But as we've seen in just a few short years, the technology has gotten better, it's gotten cheaper, and it's gotten much more accessible, making it very easy to make a realistic deepfake in a couple of minutes on your phone. As a result, we're seeing AI used to undermine elections across the world and here in America. As technology changes, so too must our laws.

Marc Berman: Therefore, I'm authoring AB 2655 to protect election integrity by regulating the online spread of AI generated disinformation and deepfakes meant to influence an election. With today's sophisticated deepfakes, voters may not know what images, audio or video they can trust. Imagine if a fake video appeared online of an elected official saying or doing something they neither said nor did, like accepting a bribe or saying that they'd hacked voting machines to ensure their victory.

Marc Berman: This deceptive but hyper realistic content can undermine voters faith in our election integrity. Accordingly, AB 2655 would, for a limited period of time immediately before and after an election, require large online platforms if they know or should know that the content meets the test in the Bill to restrict the distribution of materially deceptive and digitally altered or created images, audio or video meant to influence the election.

Marc Berman: For less harmful yet still materially deceptive and modified content, the Bill would require the platforms to label other AI created or modified election disinformation meant to influence. In both instances, the Bill does not demand perfection. What it says is that platforms must act if they know, or should know, using best available tools, that the content meets the test in the Bill. The Bill doesn't allow platforms to bury their head in the sand, but if they genuinely don't know, then there is no obligation.

Marc Berman: Moreover, the Bill is not trying to address hot button controversies or inflammatory claims, just the depiction of demonstrably untrue and provably false content. Importantly, the Bill does not provide for any monetary penalty against online platforms that fail to comply with the Bill's requirements. Instead, the only

penalty is that a court will order the platforms to do what the Bill requires.

Marc Berman: I realize that this Bill is attempting to legislate in an arena where technology and the law are fast evolving, and that can be challenging on many levels. But I don't believe that we can afford to take a wait and see approach. I've had good conversations with platforms. I've amended the Bill based on feedback, and I'll continue to do so should the Bill move forward today. Therefore, I respectfully ask for an aye vote and I'm joined today by Leora Gershenzon, CITED's Policy Director, and David Evan Harris, a lecturer at UC Berkeley and former tech sector researcher.

Tom Lackey: You may proceed.

Leora Gershenzon: Thank you, Mister Vice Chair and Members, I'm Leora Gershenzon. I'm the Policy Director at the California Initiative for Technology and Democracy, which is a project of California common cause. As elected officials, you are all keenly aware of the threat posed by deepfakes. Consider a deepfake of you hugging the opposing party's presidential candidate that goes viral weeks before an election, or a fake video of you accepting a bribe that's forwarded around your district days before the election.

Leora Gershenzon: While disinformation has been around forever, AI generated deepfakes can now be created virtually instantly, at no cost, and sent to millions in a matter of seconds. Once disinformation spreads, it's nearly impossible to get it out once it's out there, label or not, it's out there. This threat is not imaginary. Generative AI has been used around the world, and even in the United States to impact elections.

Leora Gershenzon: Obviously, the infamous robocall. AB 2655 seeks to strike the right balance by seeking to ban, only for a strictly limited time around elections, the online spread of the worst of the deepfakes. That's the candidate doing or saying something they did not do or say, or an elected official doing or something doing or saying something they did not do. Same thing with an elections official. Again, election officials a day before the election saying you can't vote. All voting machines are hacked.

Leora Gershenzon: This would obviously have a clear impact on an election and is demonstrably false. Assembly Berman has worked diligently to ensure that AB 2655's approach is very narrowly tailored and does not extend to hot punted controversies or things that are not demonstrably false. It's not a matter of opinion. It either happened or it didn't happen. The Bill is respectful of the First Amendment.

Leora Gershenzon: It's narrowly tailored to serve a compelling government interest, and it also understands the reach of Section 230 of the Federal Communications Decency Act. Will this Bill stop all election disinformation? No, but it's part of a multipronged approach to address what clearly is a crisis in our democracy. And the alternative, waiting for perfection, is simply not enough that we should accept to protect our democracy and our free and fair elections. We hope you can support AB 2655. Thank you.

David Harris: Thank you Mister Vice Chair and Members, my name is David Harris and I am a Senior Policy Advisor to CITED. I previously worked for close to five years at Facebook and Meta on the civic integrity, misinformation and responsible AI teams. I am also an advisor to a number of different organizations.

David Harris: Beyond CITED I have advised the European Union on the EU AI Act, I have advised the White House on the White House's Executive Order on AI, and I'm a Member of a NATO task force on AI and disinformation. On Assemblymember Berman's Bill 2655 I am strongly in favor of this Bill. This Bill only applies to the largest online platforms, such as the ones operated by my former employer, that have wonderful resources inside of their companies.

David Harris: To enforce a Bill like this. You only need to look to other categories of currently illegal content, such as child sexual abuse material or terrorist content, which they are able to remove very effectively. And those are not necessarily more easier to remove or simpler than election disinformation. The notion that it's too difficult to remove this is false. Beyond that, I think it's important for us to remember that most of the major AI companies today have already asked for more regulation of AI.

David Harris: And I think that is an important sign that we know that the industry is strongly in favor of putting regulations in place that protect our elections and protect our democracies. When we fail to regulate AI, what we do is we punish the companies that are doing the right thing thing. We punish them because they are going out of their way to spend more money to hire more staff and to build teams and technologies that are able to remove this type of content.

David Harris: Their shareholders who want them to maximize their profits are not pleased if they are the only companies in the space that are spending the money to do the right thing. And it's for that reason that I think we should heed the call of tech companies CEO's calling for regulation of AI, and that I am strongly in support of this Bill.

Tom Lackey: Okay, thank you. Is there anybody that would like to testify in opposition before we move forward with the group opposition. Please come forward to the table here, and you'll have five minutes to express your opposition. You may proceed.

Khara Boender: Thank you. Good morning, Chair Pellerin and Members of the Committee. My name is Khara Boender, testifying on behalf of the Computer and Communications Industry Association in respectful opposition to AB 2655. CCIA is an international, not for profit trade association with about two dozen members from a range of communications and technology firms. CCIA and its members take seriously the impact deceptive content may have on elections.

Khara Boender: Many of our members are working to implement tools to better detect and label AI generated content. And using a combination of AI and human review, they moderate content in violation of their terms or service, including content that is illegal and potentially harmful. But the tools that are currently available are not always reliable, are accurate. Just as spam filters sometimes mislabel legitimate

emails, and while such technology is evolving, so are the means for bad actors to evade such detection.

Khara Boender: I would say that this differs from the examples that were shared earlier regarding CSAM because there are databases that have hashing values that allow platforms to better detect these materials in addition to what they used to detect copyright infringements, which are again based on robust databases of available information. Because covered platforms are not privy to the intent and context for which a piece of content is used, they could inadvertently overblock or over label content. This could result in user frustration and suppression of political speech.

Khara Boender: Political speech was at the core of why our First Amendment was established, and it is critical to maintain those protections. And while the Bill exempts satire and parody, it's unclear who gets to decide what constitutes those uses. Faced with potential liability, covered platforms may choose to prohibit all digitally altered content, cutting off many valuable and helpful uses.

Khara Boender: These tools can be used by campaigns to reach voters with high quality content at lower costs or to translate speech into multiple languages to foster a more accessible democratic process. Further, users may weaponize the required reporting mechanism to suppress speech that they disagree with. For example, under the Digital Millennium Copyright Act, there were studies describing political ad takedowns on both sides of the political spectrum.

Khara Boender: After receiving reports of copyright infringement, many of these cases were ultimately deemed fair use, but platforms were inclined to err in taking down the content lest they face potential liability. Responsibility for labeling AI generated election content and liability for deceptive content should rest with the entity that puts forth such material, the entity that is most aware of the intent and context for which the content was created and shared. On behalf of CCIA, I appreciate your consideration of our comments and I welcome anyone questions.

Tracy Rosenberg: Okay. Hi once again, Committee Tracy with Oakland privacy. We are in respectful opposition to AB 2655. And we wanted to say a couple of things from a civil society, from a civil society perspective. We're not a tech company and we don't work for them or speak for them. I also want to say to Assemblymember Berman, when I was talking about the full spectrum of AI Bills, and that some were extremely complex, I honestly was talking about all the Bills, including algorithms and risk assessments and six different water marking Bills and so on, not simply this Bill. I didn't want you to take that the wrong way. So this Bill charges large online platforms with scanning all content posted to their sites with deepfake detection tools for four months preceding a vote.

Tracy Rosenberg: So if we include primaries, that's probably 60% to 75% of the time. Because we have elections every year, sometimes twice a year. We do think that is a bit of a privacy problem. It's also questionable how well these tools work. Just last week, a prominent disinformation expert said in the New York Times, even using the best tools, you can't be sure. And he was making those tools as he said that so he would know.

Tracy Rosenberg: The Bill's language sort of offers that a technology company should be the judge, jury and executioner, despite no formal training in First Amendment and constitutional law. This stuff is always a sort of, you know, push me, pull you in terms of balancing the equities. And with all due respect, we don't think TikTok or Meta would balance them as well as a judge, and we don't really want you doing that. The Bill is, you know, so fundamental.

Tracy Rosenberg: So fundamentally, the Bill is relying on two imprecise measures, technically, you know, technical scanning programs that don't necessarily work very well in unvetted reports from the public, candidates, officials, campaigns, and even chaos actors. We don't think any technology platform can be expected to know everything that every candidate running for office in every California, city, county, and, you know, across the state where they said and what they went.

Tracy Rosenberg: I mean, there's some obvious examples that were provided, but in a lot of local races, this is going to be something fairly subtle. And we don't, you know, or it can be fairly subtle, and we don't necessarily believe that TikTok, Meta or Instagram will actually know what was said and where candidates were. So basically were using imprecise measures to power a potentially broad censorship regime of blocking content. And we really can't support that even under the guise of defending democracy.

Tracy Rosenberg: So we are recommending that the Committee and the Legislature not go down this particular route. We think there are a lot of things that the Legislature can do that stop somewhat short of a broad censorship regime for online content. Thank you.

Tom Lackey: Thank you. Okay, this is now the opportunity for those in the audience who would like to express support. If you could express your name and organization, we'd love to hear from you.

Louise Miller: Louise Miller, representing the California Clean Money Campaign and Indivisible California State Strong in strong support.

Evan Minton: Hi, Evan Minton with Voices for Progress standing in strong support of this critical Bill.

Obed Franco: Good morning. Obed Franco, on behalf of the Asian Law Caucus, in support.

Tom Lackey: Thank you. Are there any in the audience who would like to express opposition? Please come forward. Express your name and organization.

Dylan Hoffman: Dylan Hoffman, on behalf of TechNet, respectfully opposed. Thanks.

Tom Lackey: Thank you. Okay, we'll now bring it back to Committee. Anybody like to express any points of clarity or - yes.

Bill Essayli: Would this cover, like, if you use chat GPT to rewrite text? Like,

does it cover text, or are we just talking about photos and videos?

Marc Berman: My understanding is it's images, audio, video, but.

Leora Gershenzon: It would cover demonstrably false text that could be told. So again, this is not a Bill that seeks perfection. And text is, I mean, you've pointed out the hardest thing to go, but can you imagine a text that goes out that says, although we don't cover web messaging, that was removed, but some content that goes out that says this person said something false. So you can. Again, it depends.

Bill Essayli: When I say text, I'm not talking about text messages.

Leora Gershenzon: No, no, I understand. You're talking.

Bill Essayli: Yes, written words.

Leora Gershenzon: Yes. But the social media platforms would have to know that it is false. And it is not a perfect science, this figuring out, is it false or is it not. It's kind of this whack a mole thing, as you expressed so well, we are getting much better with tools to determine what's false, and then the bad actors are going to get better. So it's going to go like this. But there are some very good tools, and they continue to get better, and they are getting better at taxpayers. The text is probably the hardest to detect.

Bill Essayli: Okay, so I'm just curious. I mean, if my opponent puts up an ad, it says, Bill Essayli supports insurrection, you know, and they used AI to generate that. And I'm gonna say I've never supported insurrection, but they're making that argument based on. So it's a very, very, very complicated area. So would that be.

Leora Gershenzon: That's tricky, because what does it mean to support insurrection? So I think. I don't think that's demonstrably false. If they have a picture of you with a pitchfork entering the Capitol on January 6, and you were not there, and you didn't do that. That's demonstrably false. That would be covered.

Bill Essayli: But you can see how this is.

Leora Gershenzon: But again, it's not an exact science, but what we're trying to do in the Bill is put very clear guardrails. We understand very well the limits of the First Amendment and the protections of the First Amendment provides. But, so that picture of you, and, you know, maybe it's cartoonish, and it's not obvious that it's you, that's one thing. But if it looks exactly like you and it goes out two days before the election, it's out. There's nothing you can do.

Bill Essayli: I recognize the concern. I do. I just. It's just a very sticky thing with the First Amendment and also with asking private companies to be the enforcer, essentially. And so I'm just going to say, for the record, I like the Twitter model where they use the community to sort of regulate information on there. And so, community notes is something they have where anyone can say, this is false. And I actually see it a lot. Even Elon Musk gets fact checked by his own community. So I tend to like that model where it's the public, it's the crowd sourcing is kind of

doing the moderating, then making an individual, company, or person the arbiter of. Of what's disinformation. So. But I appreciate it. Madam Chair. That's all I had. Thank you.

Gail Pellerin: Thank you. Any other comments? Assemblymember Cervantes.

Sabrina Cervantes: I'm gonna. Excuse me. So I wanna just first start off. By thanking the author and those who provided testimony today. I believe that we do need to have guardrails and strike a balance as best as we can to stop the. Spread of deepfakes and misinformation, disinformation in our elections, and just to protect our democracy. So today, I'll be supporting this Bill.

Marc Berman: Thank you.

Gail Pellerin: Any other comments? Assemblymember Lackey.

Tom Lackey: Yeah. I do clearly believe that deepfakes are a threat to truth. However, what keeps somebody from suing a website or platform in its current situation without this Bill.

Marc Berman: I'm not aware of any legal remedy or process for somebody to be able to do that. So I think that's where this Bill is sort of novel and tries to create an avenue for folks to be able to seek that redress.

Tom Lackey: And just so you know, I'm not in a position to actually support this Bill at this point just because I have my colleague has expressed some concerns that I do share, and I think it's a little too nuanced, and I think it jeopardizes something that we hold very sacred.

Marc Berman: Yeah, no, I'll wait.

Gail Pellerin: Any other comments from. Okay. Seeing no other comments. Seeing no other comments. Assemblymember Berman you may close.

Marc Berman: Yeah, I appreciate the conversation. I also appreciate how complex this is, and I don't by any stretch think that we have the perfect solution today. I think that's something that we're going to hope to try to achieve by the end of the legislative process. And even then, I have no doubt that something like this will probably be litigated in the courts.

Marc Berman: And there's a lot of interesting precedent, different opinions, and it'll be curious to see how, like I mentioned in my opening comments, this is ever evolving. And as technology changes and frankly, as certain Supreme Court Justices change, we don't know how the Supreme Court might opine on the influence of technology and manipulated content and the impact that it has on our elections and what scrutiny they'd apply and all sorts of things.

Marc Berman: So it's a really fascinating Bill, probably one of the more challenging ones I've got this year. And it's something that we're going to keep on having conversations with the opposition to try to make sure that we are addressing the

concerns that we have, or that I have about the impact of this type of content on our elections and on our voters without unintended consequences, and doing it in a way that's as easy as possible for the platforms to implement. So with that respect, we ask for an aye vote.

Gail Pellerin: Thank you so much, and I do appreciate your leadership in this issue. You were out there early on, the first in the nation. I really want to thank you for that foresight. The threat posed to our electoral system and to our democracy by deepfakes and other misleading content is substantial and has been supercharged by tools like artificial intelligence that make it easier and cheaper than ever to create convincingly realistic but deceptive images, audio and video.

Gail Pellerin: And part of the solution for protecting against the deceptive material is to go after those who create and disseminate that material, which is why, which is why you're doing this Bill. So thank you. But another crucial piece of the puzzle is to address the harms that these materials present is to limit the tools that bad actors have for spreading disinformation so rapidly and at such little cost.

Gail Pellerin: So I know that the author takes all this seriously, the constitutional and logistical challenges and regulating in this area, and I'm confident that you're open to continuing to refine the Bill as it moves through the process. So with that, I'm recommending support. Do I have a motion? Motion? Motion by Cervantes, second by me. Madam Secretary, please call the roll

Committee Secretary: On AB 2655 by Berman. The motion is do pass and be rereferred to the Committee on Judiciary. [Roll Call].

Gail Pellerin: We're going to keep that Bill on call and we're going to round up our absent Members. So thank you. We'll now move on to Assemblymember Cervantes and your Bill 1807.

Exhibit 8

Senate Standing Committee on Judiciary
AB 2655

Thomas Umberg: Put that on call. Thank you very much. Thank you, sir. All right, summary Member Berman, I think I saw you here. Okay, Senator. Member Berman, filing number 60, AB 2655.

Marc Berman: Thank you, Senator Umberg. And, Senator, I'd like to start by thanking the Committee consultant for his thorough analysis and for taking the time to work with my office while juggling the overwhelming workload for this hearing. I'm happy to take the amendments outlined in the analysis. Five years ago, I authored the first election deepfake Bill in the nation.

Marc Berman: Just a few short years later, the technology is better, cheaper, and more widely accessible. And we're seeing deepfakes used to undermine elections across the globe and here in America.

Marc Berman: This past month, actually, a study from Google's DeepMind AI division confirmed our fear finding that AI generated deepfakes that impersonate politicians and celebrities are the most popular ways to misuse AI. The report specifically found users deploy a range of tactics to distort the public's perception of political realities.

Marc Berman: That's why I'm authoring AB 2655 to protect election integrity by requiring large online platforms to, for a limited time, restrict the distribution of materially deceptive content intended to sway voters or undermine confidence in elections. For less harmful, yet still materially deceptive content, the Bill would require the platforms to label it as election disinformation.

Marc Berman: I've got a couple other comments, but I'll save them for later. Respectfully ask for aye vote and with me today are two fantastic witnesses on behalf of the sponsor cited who can go into a little more detail. All right, on the Bill, first fantastic witness. Go ahead.

Leora Gershenzon: That's tough to live up to. Chair and Members, I'm Leora Gershenzon with cited the California Initiative for Technology and Democracy, a project of California common cause. AB 2655 seeks to protect the integrity of our elections from the worst of the worst deepfakes meant to defraud our electorate and undermine faith in our democracy.

Leora Gershenzon: We believe that AB 2655 strikes the right balance by banning, for a strictly limited time around the elections, the online spread of the worst deepfakes intended to deceive voters and influence elections.

Leora Gershenzon: Videos, images, and audios of candidates, election officials, and elected officials doing or saying something they didn't do or say that appear real and that are reasonably likely to change the election outcomes or falsely undermine confidence in the election.

Leora Gershenzon: The Bill also requires that other fake online content that isn't

quite as bad is labeled again for a limited time around the election. While this Bill can't stop all election misinformation, it will help remove some of the worst deepfakes from the ecosystem close to election times when voters are paying the most attention.

Leora Gershenzon: Please help us protect our democracy. Thank you, alrighty.

David Harris: Thank you. Next fantastic witness. Thank you, chair Umberg. It's an honor to be here with you today. My name is David Harris and I'm a senior policy advisor to CITED. I also teach at UC Berkeley at the Haas School of Business about AI ethics and policy, and I've been an advisor to numerous international bodies.

David Harris: I've advised the European Union, the White House, NATO, all on AI policy. I'm here in strong support of AB 2655 because it takes action where tech companies will not.

David Harris: I previously worked for close to five years at Facebook and Meta where I was a research manager on the responsible AI team and also on the civic integrity team. In those companies, we have a problem, which is that the companies are not willing to do anything that is not legally required of them.

David Harris: Where deepfakes and elections and AI are concerned, this is difficult because what it means is that when certain companies step out and take responsibility for these problems, they get punished.

David Harris: They get punished by their shareholders and their constituents and they have to lay off the workers that are responsible for detecting deep fakes and getting rid of these kinds of efforts at election interference. An interesting thing that perhaps makes the AI industry unique is that many AI CEO's have themselves demanded more regulation of artificial intelligence.

David Harris: And that's why bills like AB 2655 that place clear rules about what AI can be used for and what social media platforms can and cannot display and distribute to their users are necessary.

David Harris: So I'm here to call on you to give the tech company CEO's exactly what they are asking for, which is better regulation of artificial intelligence where our democracy is concerned.

Thomas Umberg: Thank you very much. Others in support, please approach the microphone. Name, affiliation and position.

Lifton Wilson: Lifton Wilson on behalf of the city and County of San Francisco Board of Supervisors in support. Thank you. Thank you.

Benjamin Cohen: Benjamin Cohen of Mountain View in support.

Thomas Umberg: Thank you. Others in support seeing no one else approached the microphone, let's turn to the opposition. If you're opposed to filing number 60, AB 2655 you may either take a chair or come to the microphone. Go ahead, whoever would

like to go first? Okay, first, fantastic witness in opposition.

Naomi Padron: Okay, perfect. Thank you, Mister chair and Members of the Committee, my name is Naomi Padrone, testifying on behalf of the Computer and Communications Industry Association in respectful opposition to AB 2655 CCIA and its Members take seriously the impact deceptive content may have on elections.

Naomi Padron: Many of our Members are working diligently to implement tools to better detect and label AI generated content. But the tools that are currently available are not always reliable or accurate, and while such technology is evolving, so are the means for bad actors to evade detection.

Naomi Padron: We appreciate the recent amendments as well as those referenced in the Committee analysis. However, we remain opposed to the Bill and are specifically concerned about the potential impacts on free speech, an issue that is also thoroughly outlined in the analysis.

Naomi Padron: Presumably, under AB 2655 the injunctive relief being sought would result in the blocking or labeling of reported content. This would be weaponized by competing parties that could simply disagree with the content or each other.

Naomi Padron: The Committee analysis notes that the Bill standards, quote, place a burden on platforms to establish facts potentially well outside of their bounds of knowing. Indeed, platforms are not privy to the intent and context for which a piece of content is used and could inadvertently overblock over a label.

Naomi Padron: This could result in user frustration and suppression of political speech. Political speech was at the core of why our first amendment was established. Responsibility for labeling AI generated election content and liability for deceptive content should rest with the entity that originally puts forth such material.

Naomi Padron: Additionally, we have concerns about the scope of the Bill while it establishes defined periods surrounding elections. AB 2655 does not speak specify which elections and where. This could result in platforms being required to block content almost constantly to ensure compliance. Thank you for your time and consideration of our comments. We respectfully request a no vote.

Thomas Umberg: All right, thank you very much. Not to be judgy. Second, fantastic witness in opposition.

Tracy Rosenberg: Thank you. Tracy Rosenberg Oakland Privacy AB 2655 sets up social media companies who, with all due respect, have no constitutional law training, no training in collections law as judge, jury and executioner. They're relying on two pretty imprecise measures, public reports from other candidates or chaos actors or scanning tools. Neither works.

Tracy Rosenberg: Deepfake detection services have been fooled into declaring images of kissing robots and giant neanderthals to be real when they're not. This is not what people want.

Tracy Rosenberg: A 2023 fire poll reported that 61% of Democrats, 62% of independents, and an even higher percentage of percentage of Republicans do not trust social media companies to moderate content on their platforms fairly. We believe that labeling synthetic content is useful and it helps out voters.

Tracy Rosenberg: But a broad censorship regime by the titans of tack is not the way that we should go, and we ask for your no vote.

Thomas Umberg: Thank you very much. Others in opposition to AB 2655-2655 thank you.

Ronak Daylami: Ronak the Alami with Cal Chamber, also opposed. Thank you.

Unidentified Speaker: Thank you. With the Electronic Frontier foundation in opposition. Thank you. Thank you.

Bryant Miramontes: Mister chair. Committee Members, Brian Marimontas of American Federation of State County Municipal Employees, apologies for being out of order. We are in support. In support.

Thomas Umberg: Okay, got it.

Cynthia Valencia: Cynthia Valencia, legislative advocate with the ACLU. California action in opposition due to first Amendment concerns.

Thomas Umberg: All right, thank you.

Ronak Daylami: I apologize also on behalf of my colleague@Tech.net. opposed. Thank you.

Thomas Umberg: Thank you. All right, bring it back. Committee questions by Committee Members, questions by Senator Stern.

Henry Stern: Yeah, thank you to the author. I'm sure this is somewhat of a reluctant effort in that wish we didn't have to be in this position to sort of push the envelope here and do what we have to do and step in, but our democracy is truly at risk.

Henry Stern: My question is, I don't trust the Supreme Court on these issues these days. And either from you or the experts, I want to see this in law. I want to see this executed.

Henry Stern: But does some of that, I forget the name of the cases, but some of the work in appellate courts, I think it's in some of the medicases, free speech. Is there ways we can prevent against that sort of falling into any traps that might occur out of some of that appellate review?

Henry Stern: I think pursuant to a Florida law, if I recall correctly, recent cases.

Marc Berman: Yeah, yeah. Let me just say at a high level, we are very eager to do

everything we can to make sure it's as narrowly tailored as possible. I think everyone agrees it's addressing a compelling government interest. So we check that box. Active litigation happening right now.

Marc Berman: I think the Supreme Court just sent those cases back to those appellate courts because they didn't think they applied the law correctly. But I'll defer to my witnesses, who might have a little more detail on that.

Leora Gershenzon: There's nothing in those cases that actually would change this Bill. Those cases addressed for speech. And in fact, there was one issue on for speech that Committee amendments suggest, Committee amendments propose we take that would limit that. So that's the only similar issue.

Leora Gershenzon: What it did say is that the court should actually consider how these First Amendment cases are applied and not just look at the face, how they are applied facially. So you have to look at them applied to the facts. Regardless, in all likelihood, this Bill would be litigated, because it's a question of how narrowly tailored are you?

Leora Gershenzon: We have a compelling state interest, which is protecting our democracy, and then is this narrowly tailored enough to serve that compelling interest? It is hard to believe, as most of the tech bills, that someone challenges them. But the point is to actually make some good law. And I don't know that it necessarily reaches the Supreme Court.

Leora Gershenzon: Some California cases reach the Supreme Court, most don't. But what we've done is try to make the strongest possible case under the First Amendment and under Section 230 for this Bill to move forward. And we think we've done that.

Marc Berman: And I want to reiterate for everybody, my office is wide open for suggestions for how to make it more narrowly tailored. I had a chat with my chief yesterday. So if folks have any ideas that don't undermine the Bill, then I'm very open to hearing those.

Henry Stern: And I would say not to force this work on our Attorney General, but I know that we're currently being challenged on a compelled political speech case for a climate disclosure law that the US Chamber and the California Chamber are challenging California's ability under the First Amendment to require companies to disclose their greenhouse gas emissions, and calling that a sort of a form of compelled political speech.

Henry Stern: This is a new and very important area of law for the Attorney General to be able to defend our interests, whether it's in our public health and safety or the integrity of our elections. So I think it is going to take some heavy lifting within our Department of Justice.

Henry Stern: So hopefully, before this hits the floor, too, just giving it that scrub and getting ourselves ready to prevail so we can actually get to an implementation stage. Anyhow, I'm happy to move the Bill. Really appreciate you just going deep on this, and hopefully we can get this into law and through our court

system.

Thomas Umberg: With everything you said. Thank you. I think. Senator Stern, you just moved the Bill. Yes, I did. Yes, you did. All right, thank you very much. Would you like to close? I ask for your aye vote okay. Thank you. Madam Chief counsel, please call a roll.

Committee Secretary: File, item number 60 AB 2655. The motion is to pass, as amended, to appropriations. [Roll Call] That's two to one with Members missing.

Thomas Umberg: Alrighty. Two to one. We'll put that on call next. Assemblymember Lowenthal, two matters, and I'm going to turn over the chairmanship here for a moment to Senator Stern.

Exhibit 9



← Post



Kamala HQ
@KamalaHQ



Vance: Democrats want to attack Republicans as being anti-union and sometimes the shoe fits



12:57 PM · Aug 29, 2024 · **159.5K** Views

708 Reposts **53** Quotes **2,723** Likes **66** Bookmarks



Don't miss what's happening

People on X are the first to know.

Log in

Sign up

New to X?

Sign up now to get your own personalized timeline!

Sign up with Google

Sign up with Apple

Create account

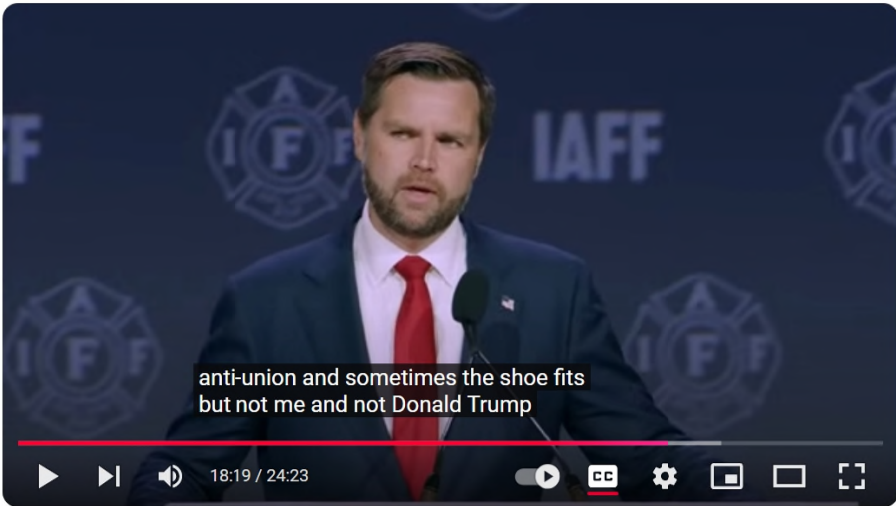
By signing up, you agree to the [Terms of Service](#) and [Privacy Policy](#), including [Cookie Use](#).

Something went wrong. Try reloading.

Retry

[Terms of Service](#) [Privacy Policy](#) [Cookie Policy](#)
[Accessibility](#) [Ads info](#) [More ...](#) © 2024 X Corp.

Exhibit 10



2024 United States elections

Get the latest information about the election outcome from the AP on Google.

[Learn more](#)

57th IAFF Convention: Sen. JD Vance



The International Ass...

10.7K subscribers

Subscribe

97



Share



5,990 views Aug 29, 2024

Republican Vice Presidential nominee and Ohio senator JD Vance addresses delegates at the IAFF's 57th Convention.

Music

1 songs



America First
Merle Haggard
Chicago Wind

Music

Transcript

Follow along using the transcript.

[Show transcript](#)



The International Association of Fire Fighters

10.7K subscribers



Videos



About



Twitter

Show less

51 Comments

Sort by



Add a comment...



2024 Buick's Are Turning Heads

The 2024 Buick SUV Lineup Is Turning Heads. Discover Sales Using Top Searches.

Sponsored · All Things Auto

[Visit site](#)



BREAKING: JD Vance Met With Mixed Reception At...

Forbes Breaking News
1.1M views · 2 months ago



J.D. Vance's wife, Usha Vance, introduces her husband at...

CNBC Television
1.1M views · 3 months ago



FULL REMARKS: Jim Gaffigan Shows No Mercy To Democrat...

Forbes Breaking News
2.5M views · 2 weeks ago



Jake Tapper and JD Vance spar over John Kelly. Watch the full...

CNN
1.3M views · 9 days ago



WATCH LIVE: Kornacki Cam on Election Night 2024 | MSNBC

MSNBC
174K watching
LIVE



Vance speaks at the International Association of...

The National Desk
4.3K views · Streamed 2 months ago



J.D. Vance addresses RNC crowd: FULL SPEECH

FOX 4 Dallas-Fort Worth
1.9M views · 3 months ago



Senator JD Vance holds campaign rally in Newtown, Pa...

FOX 29 Philadelphia
16K views · Streamed 1 month ago



JD Vance Exclusive Interview

wgaltv
10K views · 1 month ago



Victor Davis Hanson on "The Case For Trump"

Hoover Institution
10M views · 5 years ago



IAFF 57th Convention: Gov. Tim Walz

The International Association of Fire...
4.1K views · 2 months ago



He Predicted Trump in 2016, Biden in 2020 and Now... | NYT...

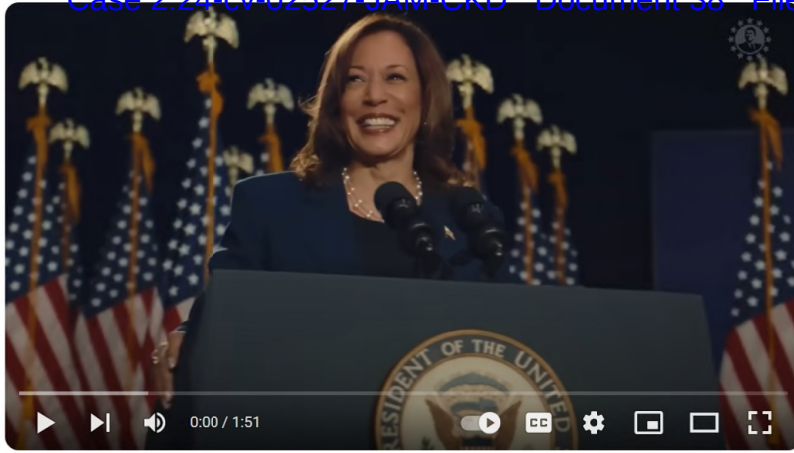
The New York Times
3.3M views · 2 months ago



LIVE: JD Vance meets firefighters union in Boston

The Times and The Sunday Tim...

Exhibit 11



2024 United States elections
Get the latest information about the election outcome from the AP on Google. [Learn more](#)

Kamala Harris Ad PARODY



Mr Reagan
383K subscribers

Subscribe

26K



Share



450,213 views Jul 26, 2024 #Politics #News #Trending
Thanks to Elon Musk for the tweet!
Kamala Harris just posted her first 2024 presidential campaign ad.
It's clean and professional. It's very well done.
So, of course, I had to produce a parody.

Patreon: [mrreagan](#)

MR REAGAN MERCHANDISE
<https://teespring.com/stores/mr-reagan>

FOLLOW MR REAGAN ON TWITTER!
[mrreaganusa](#)

Music by The Passion HiFi
www.thepassionhifi.com

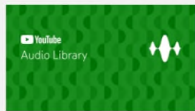
#Politics #News #Trending

How this content was made

Altered or synthetic content
Sound or visuals were significantly edited or digitally generated. [Learn more](#)

Music

1 songs



Greengrass
Drew Banga
Greengrass

Music

Transcript

Follow along using the transcript.

Show transcript



Mr Reagan
383K subscribers

Videos

About

Twitter

Patreon

Facebook

Show less



Play Free On Desktop
Boldly Go Where No One Has Gone Before
Sponsored · Scopely

Download



The Best of President Reagan's Humor
Ronald Reagan Presidential Foundat...
9M views · 7 years ago



Brett Favre Tees Off On Biden For Calling Trump Supporters...
Forbes Breaking News
1.1M views · 6 days ago
New



What's That Name: Election Edition - SNL
Saturday Night Live
1.8M views · 2 days ago
New



Some early states called for Trump, Harris on election night
CBS Chicago
4.4K views · 1 hour ago
New



Fox News Kamala Harris Interview Cold Open - SNL
Saturday Night Live
8M views · 2 weeks ago



Presidential NIGHT LIVE UPDATE AND LATEST...
H. A. Goodman
793 watching
(LIVE)



Ronald Reagan Sits Down with Johnny | Carson Tonight Show
Johnny Carson
5M views · 4 years ago



Family Feud Election 2024 Cold Open - SNL
Saturday Night Live
10M views · 3 weeks ago



Donald Trump on Letterman, May 21, 1992
Don Giller
3.3M views · 8 years ago



Live 2024 Election Results and News Coverage | FOX 5 New...
FOX 5 New York
1.2K watching
(LIVE)



Why 27 U.S. States Are Going Broke
CNBC
376K views · 1 day ago
New



Hysterical Kamala Harris parody ad goes viral after being bann...
Sky News Australia
7M views · 1 month ago



Jim Justice flips West Virginia Senate seat
FOX 5 Washington DC
9.1K views · 44 minutes ago
New



Man Identifying As 6-Year-Old Breaks All Records In T-Ball...
The Babylon Bee
476K views · 1 month ago



Maddow: Despite right-wing threats, pro-Trump election...
MSNBC
163K views · 19 hours ago
New



Best Fails of the Year (So Far)

Exhibit 12



← Post




Elon Musk  
@elonmusk



This is amazing 🤔



From **Mr Reagan** 

7:11 PM · Jul 26, 2024 · **136.6M** Views

216.2K Reposts **26.5K** Quotes **920.5K** Likes **105.5K** Bookmarks





105K



New to X?

Sign up now to get your own personalized timeline!

 Sign up with Google

 Sign up with Apple

Create account

By signing up, you agree to the [Terms of Service](#) and [Privacy Policy](#), including [Cookie Use](#).

Something went wrong. Try reloading.

 Retry

[Terms of Service](#) [Privacy Policy](#) [Cookie Policy](#)
[Accessibility](#) [Ads info](#) [More ...](#) © 2024 X Corp.

Don't miss what's happening

People on X are the first to know.

Log in

Sign up

Exhibit 13

[← Post](#)

Gavin Newsom
@GavinNewsom



I just signed a bill to make this illegal in the state of California.

You can no longer knowingly distribute an ad or other election communications that contain materially deceptive content -- including deepfakes.



Gavin Newsom @GavinNewsom · Jul 28

Manipulating a voice in an "ad" like this one should be illegal.

I'll be signing a bill in a matter of weeks to make sure it is.



Elon Musk retweets altered Kamala Harris campaign ad

In the video, Harris seemingly exposes herself as an incompetent candidate for president. The origin of the video isn't known yet.

7:41 PM · Sep 17, 2024 · **25.9M** Views

24.5K Reposts **5,295** Quotes **127.6K** Likes **4,003** Bookmarks



4K



New to X?

Sign up now to get your own personalized timeline!



Sign up with Google



Sign up with Apple

Create account

By signing up, you agree to the [Terms of Service](#) and [Privacy Policy](#), including [Cookie Use](#).

Something went wrong. Try reloading.



Retry

[Terms of Service](#) [Privacy Policy](#) [Cookie Policy](#)
[Accessibility](#) [Ads info](#) [More ...](#) © 2024 X Corp.

Don't miss what's happening

People on X are the first to know.

Log in

Sign up

Exhibit 14

[Hearings](#)

Share

Senate Standing Committee on Elections and Constitutional Amendments

June 18, 2024

[Return to comment starting point](#)

Search transcript

Khara Boender
Person

While establishes defined time periods surrounding elections and election processes for newly established prohibitions and requirements, the Bill does not specify which elections where. As noted in the Committee analysis, this could result in platforms being required to block content almost constantly in order to ensure compliance.

Khara Boender
Person

We appreciate the recent amendments to limit the PRA to a candidate for elective office, elected official or elections official in lieu of any resident of California. However, we're still concerned that this could still result in the weaponization of political speech that competing candidates disagree with.

Khara Boender
Person

For example, under the Digital Millennium Copyright act, there were studies describing political ad takedowns on both sides of the political spectrum. After receiving reports of copyright infringement. Many of these cases were ultimately deemed fair use, but platforms were inclined to err in taking down the content lest they faced liability.

Khara Boender
Person

For these reasons, we urge a no vote on this legislation. On behalf of CCIA, I appreciate your consideration. Thank you.

Catherine Blakespear
Legislator

Thank you very much. Any other opposition witnesses in the room?

Ruth Dawson
Person

Good morning. Ruth Dawson with ACLU, California Action and respectful opposition thank you.

View Agenda

CURRENTLY DISCUSSING

Bill AB 2655

Defending Democracy from Deepfake Deception Act of 2024.

[View Bill Detail](#)[Download Transcript](#)

COMMITTEE ACTION: PASSED ✓

FOR 6 AGAINST 1 NO VOTE 0

[Benjamin Allen](#)[Catherine Blakespear](#)[Anthony Portantino](#)[Josh Newman](#)[Show all 6 votes](#)Next bill discussion: [July 2, 2024](#)Previous bill discussion: [May 22, 2024](#)

Speakers

LEGISLATOR

[Marc Berman](#)[Catherine Blakespear](#)[Josh Newman](#)

ADVOCATE

[Khara Boender](#)[Dylan Elliott](#)[David Harris](#)[Trent Lange](#)[Show more speakers](#)

Journalists backed by artificial intelligence bringing transparency and accountability to California's policy choices.

[About Digital Democracy](#)[Data & Methodology](#)[Visit CalMatters](#)

Support this nonprofit initiative

Digital Democracy informs Californians, holds officials accountable, and builds a stronger community. Brought to you by the 501(c)(3) nonprofit and nonpartisan CalMatters newsroom.

[Support Our Work](#)

Sign up for news and updates

Receive updates about Digital Democracy and CalMatters' daily newsletter that brings transparency to state government.

Email

[Sign Up](#)By signing up, you agree to the [terms](#).

Senate Standing Committee on Elections and Constitutional Amendments
AB 2655

Marc Berman: Thank you, Chair Blakespear and Sanders. Five years ago, I authored the first election deepfake Bill in the nation. And as you heard with the presentation of a Bill earlier, just a few short years later, the technology is better, cheaper, and more widely accessible.

Catherine Blakespear: AB 2655.

Marc Berman: And we're seeing deepfakes used to undermine elections across the globe and here in America.

Marc Berman: Therefore, I'm authoring AB 2655 to protect election integrity by requiring large online platforms to, for a limited time, restrict the distribution of materially deceptive content intended to affect the outcome of an election, undermine confidence in election results, or harm the reputation of a candidate for less harmful, yet still materially deceptive content.

Marc Berman: The Bill would require the platforms to label it as election disinformation. With today's sophisticated deepfakes, voters may not know what audio or visual content they can trust, which can undermine election integrity.

Marc Berman: Imagine if a fake video appeared online of an elected official doing or saying something that they did not do or say, like accepting a bribe or saying that they'd hacked voting machines to ensure their own victory. The Bill does not demand perfection. What it says is that platforms must act using the best available tools.

Marc Berman: If they know or should know that the content meets the test in the Bill platforms cannot bury their head in the sand, but if they don't know, then there is no obligation.

Marc Berman: I recognize that this Bill is legislating in an arena where technology and the law are fast evolving, and this can be challenging, but doing nothing is not an option. I respectfully ask for your aye vote. And with me today are Laura Gershenzon and David Harris, both with the California Institute for Technology and Democracy.

Catherine Blakespear: Thank you and welcome. You each have two minutes.

Leora Gershenzon: Thank you. Leora Gershenzon, we're CITED the California Initiative for Technology and Democracy, which is a project of California Common Cause. Imagine two different videos that are going viral online. One is an influencer on TikTok saying, don't vote by mail. It's bad. You may not get your ballot counted. Go in person.

Leora Gershenzon: The other is the election official in your county, a deep fake of them saying don't vote by mail. We throw them all out. They're two fundamentally different things. Both are designed to impact the vote and impact two votes. But one you can counter with more information you can have saying no, come and vote.

Leora Gershenzon: The other one is just so bad, so fundamentally bad, that having it out there, even if it had a line on the bottom that said, this is AI, is simply not enough, that it just shouldn't be in the ecosystem, because once it's out there, it is impossible to take back.

Leora Gershenzon: And this Bill is trying to do that, is trying to say the worst of the worst election Deepfakes.

Leora Gershenzon: Those are the candidate or the elections official or an elected official doing or saying something they didn't do or say with the intent to influence the election that comes up just shouldn't be there for a short time frame around the election. And other things in a longer timeframe should be labeled as such.

Leora Gershenzon: The bottom line is there are just things that cannot be unseen and will impact our democracy and whether people vote or not. And we hope and pray that this is enough to be able to. This Bill and the other bills that you're hearing are enough to protect California's democracy. We urge you to vote aye.

Catherine Blakespear: Thank you very much. Yes, go ahead.

David Harris: Chair Blakespear and Members of the Committee, it's an honor to be here speaking with you. My name is David Harris, and I'm a senior policy advisor to cited the California California Initiative for Technology and Democracy.

David Harris: I am also chancellor's public scholar at UC Berkeley, where I teach classes at the Haas School of Business on topics including AI, ethics for leaders, and civic technology. I previously worked for close to five years at Facebook and Meta on the civic integrity and responsible AI teams.

David Harris: I've also advised the EU, the White House, NATO, and the United nations on artificial intelligence. This Bill only applies to the largest online platforms with the greatest reach of potential election misinformation, and it's fully implementable today based on tools the companies already possess. Illegal content is complex, child sexual abuse material, terrorist content.

David Harris: This Bill is very narrow and makes it impermissible to generate election disinformation. The companies have already shown that they can do it with those other types of content.

David Harris: Some of these companies suggest that the requirements cannot be met, but they already are under similar requirements in Europe due to Europe's Digital Services act, which is designed to, among other things, crack down on election interference. So they are already complying there or trying to.

David Harris: They've also promised to do these same types of reasonable things that the Bill provides. The tech companies made a big commitment in the AI and elections accord at the Munich Security conference this past February.

David Harris: But unfortunately, the commitments that the tech companies have made in places like that are voluntary, have no timelines for implementation, and have no accountability mechanisms associated with them. The good news though is that AI companies and social media companies have been publicly calling for more regulation of their industry.

David Harris: And regulation is great in this field because it creates a level playing field where all the companies have to play by the same rules. Without regulations like this, you punish the companies that are working hard to operate in the interests of society.

David Harris: And I'll close by saying that we should heed the call of tech industry CEO's calling for regulation and promising to do better for democracy and give them what they are asking for with AB 2655.

Catherine Blakespear: Thank you very much for your testimony. Do we have any other support in the room? Please come forward to the microphone.

Dora Rose: Morning, Dora Rose, League of Women Voters of California in strong support and I have the proxy of Disability Rights California also in strong support. Thank you.

Janice O'Malley: Chair, Members, Janice O'Malley AFSME California in support thank you.

Dylan Elliott: Dylan Elliott, on behalf of the San Francisco Board of Supervisors in support. Thank you.

Evan Minton: Hi, Evan Minton, Voices for Progress in support thank you.

Marquise Mason: Marquise Mason, California Environment of Voters in support thanks.

Trent Lange: Trent Lange, California Clean Money Campaign in support.

Catherine Blakespear: Okay, thank you very much. Do we have any lead witnesses in opposition.

Catherine Blakespear: Hello. Welcome. Come forward. You may start when ready. You have two minutes.

Khara Boender: Great. Thank you. Madam Chair, Members of the Committee, again, my name is Khara Boender. On behalf of the Computer and Communications Industry Association, in respectful opposition to AB 2655 CCIA and its Members take seriously the impact deceptive content may have on elections.

Khara Boender: Many of our Members, as noted earlier, are working to implement tools to better detect and label AI generated content and using a combination of AI and human review. They moderate content in violation of their terms of service, including content that is illegal and potentially harmful.

Khara Boender: But the tools that are currently available are not always reliable or accurate, and while such technology is evolving, so are the means for bad actors to evade such detection. Because covered platforms are not privy to the intent and context for which a piece of content is used, they could inadvertently overblock or over label content.

Khara Boender: This could result in user frustration and suppression of political speech. Political speech was at the core of why our first amendment was established.

Khara Boender: Therefore, responsibility for labeling AI generated election content and liability for deceptive content should rest with the entity that puts forth such material, the one that is most aware of the intent and context to which the content was created and shared. We also have concerns about the breadth of the Bill.

Khara Boender: While establishes defined time periods surrounding elections and election processes for newly established prohibitions and requirements, the Bill does not specify which elections where. As noted in the Committee analysis, this could result in platforms being required to block content almost constantly in order to ensure compliance.

Khara Boender: We appreciate the recent amendments to limit the PRA to a candidate for elective office, elected official or elections official in lieu of any resident of California. However, we're still concerned that this could still result in the weaponization of political speech that competing candidates disagree with.

Khara Boender: For example, under the Digital Millennium Copyright act, there were studies describing political ad takedowns on both sides of the political spectrum. After receiving reports of copyright infringement. Many of these cases were ultimately deemed fair use, but platforms were inclined to err in taking down the content lest they faced liability.

Khara Boender: For these reasons, we urge a no vote on this legislation. On behalf of CCIA, I appreciate your consideration. Thank you.

Catherine Blakespear: Thank you very much. Any other opposition witnesses in the room?

Ruth Dawson: Good morning. Ruth Dawson with ACLU, California Action and respectful opposition thank you.

Jose Torres Casillas : Jose Torres Casillas, on behalf of Technet and Chamber of Commerce in respect for opposition.

Brandon Knapp: Brandon Knapp, representing Electronic Frontier foundation, respectfully in opposition.

Catherine Blakespear: Thank you very much and we will now bring it back to our Members. And I'd like to ask the author, would you like to respond to any of the opposition witnesses statements?

Marc Berman: Appreciate the concerns. We've been working together, having a lot of conversations with all the stakeholders since the day we introduced the Bill. You know, we believe that we have narrowly tailored the Bill to address a significant government interest. Also believe that we've put language in there to make sure that the platforms aren't held to perfect standard.

Marc Berman: But for the big platforms that the Bill applies to, that they are using best practices that, as one of my witnesses mentioned, they're already using for other users in other parts of the world.

Marc Berman: But at the same time, we'll absolutely continue having conversations with the opposition and looking for ways to address concerns that they have that don't totally undermine or gut the purpose of the Bill. And I'm sure my witnesses, if given the opportunity, might also have some thoughts. But through the chair.

Catherine Blakespear: I mean, if you'd like to, briefly, I just wanted to give you a chance.

Unidentified Speaker: The only thing I would add is the liability under this Bill is just injunctive relief. So all it is is a court telling the platform to take down what they should have already taken down. There is not monetary liability, liability, thanks to Section 230.

Unidentified Speaker: I'd just like to respond respectfully to the point raised by the opposition that the tools are not reliable or accurate.

Unidentified Speaker: There are well established tools for doing this, and that's why when you open up your Facebook, newsfeed, or your Instagram feed, you generally will not find pornography, child sexual abuse, terrorist recruiting content, because they can remove that type of content.

Unidentified Speaker: There's no fundamental difference between that type of content and this type of content, except that this type of content is not yet illegal, and that's why the companies are not doing it. They will do it if forced to. Elon Musk got rid of about 81% of Twitter's staff after he acquired it.

Unidentified Speaker: Mark Zuckerberg said publicly that he admired what a lean company Elon had been able to make Twitter. And that's because activist investors are saying, cut, cut, cut, make your companies more profitable. Only do what you're legally required to do.

Unidentified Speaker: And it is on us and you Members of the Committee, to make it legally required to address this issue.

Catherine Blakespear: Okay, well, thank you very much. I appreciate that, Senator Newman.

Josh Newman: Thank you. And so I guess I'd like to ask a question. The opposition. So, Miss Punder, you happen to be at the table, so apologies. You're representing all of those nice folks, but would you agree or disagree that this is, in fact, a

fairly urgent matter, given the advances in AI and the prospective threat to our elections.

Khara Boender: That it's important to address the impacts that AI may have on elections. But I believe that is why we are supportive of the other measures that put the responsibility on the people who are producing the content that is in fact deceptive or they know how they have produced it and in what context that they have produced it.

Khara Boender: To back up a little bit on that front, while respectfully, the support witness over here mentioned that the reason why when you open up Facebook and you're not seeing a bunch of CSAM, I would argue that that's a fundamentally different and not apples to apples comparison as CSAM is identified using hashing values that are found at a database, which is not the case for something that may be considered more subjective when we're talking about whether a specific person said or did something.

Khara Boender: And I think that's more difficult for a platform to know definitively whether somebody, you know, across the country, there are over 7000 state legislators in the country, whether that specific candidate did say or in fact do something. So I do think that those are fundamentally different examples that are being shared.

Josh Newman: I appreciate that. But I guess the use case that the author is trying to address, you know, I regard as a real problem and at least visually not hard to identify once identified to the support positions point. It's really a work factor question. Right.

Josh Newman: And so the question at issue here is, you know, what's the willingness of the major platforms to do this work? And in the absence of this legislation, how can we be assured as representatives of the public interest that you're going to do that work?

Josh Newman: So this seems in the absence of an alternative, I'd be interested to hear an alternative solution. Seems like a very reasonable measure. So do you have, is there a better way to do this that gets at the issue? Because I think the issue is fairly clear cut.

Khara Boender: Again, I think it's about going after the people who are producing the material originally.

Josh Newman: Okay. That is a whack a mole Proposition unless we do that at a systems level. Okay. Clay was interested to hear but happy to support the Bill in its current form. Thank you.

Catherine Blakespear: Okay. Thank you. Assembly Member Berman, would you like to close?

Marc Berman: I appreciate the conversation. We'll continue to work with the opposition. Respectfully ask for your aye vote.

Catherine Blakespear: Thank you. Do we have a motion? Okay, Senator Newman, thank you. This is due past two. Judiciary assistant, please call the roll.

Committee Secretary: Senators Blakespear? Blakespear aye. Nguyen. Nguyen no. Allen. Allen aye. Menjivar aye. Newman. Newman aye. Portantino. Umberg.

Catherine Blakespear: It's 41 and we will leave it on call. Thank you.

Marc Berman: Thank you.

Catherine Blakespear: Thank you very much. So now we are on to our last item of the day here for these Elections and Constitutional Amendments Committee. It's ACA 8. We will ask Assembly Member Wilson to come forward. You may proceed when ready.

Exhibit 15

Harvard Journal of Law & Technology
Volume 24, Number 1 Fall 2010

**FREE SPEECH UNMOORED IN COPYRIGHT’S SAFE HARBOR:
CHILLING EFFECTS OF THE DMCA ON THE FIRST
AMENDMENT**

*Wendy Seltzer**

TABLE OF CONTENTS

I. INTRODUCTION.....	171
II. DMCA’S CHILL AS “PRIOR RESTRAINT BY PROXY”	177
A. <i>Error Costs of the DMCA</i>	179
B. <i>Intermediation</i>	181
C. <i>The DMCA and the Economics of Speech</i>	184
III. FIRST AMENDMENT PROBLEMS	187
A. <i>Copyright and the First Amendment</i>	187
B. <i>Prior Restraints on Speech</i>	190
C. <i>Understanding Chilling Effects</i>	193
D. <i>Chill, Intermediated</i>	197
IV. THE CHILL WINDS OF COPYRIGHT AND DMCA.....	200
A. <i>Errors and pressures</i>	204
B. <i>The Chill in Practice</i>	210
C. <i>“Repeat Infringers”</i>	218
D. <i>Limited Warming?</i>	221
E. <i>Against Copyright Secondary Liability</i>	225
V. REFORMING COPYRIGHT TAKEDOWN.....	226

I. INTRODUCTION

The 2008 U.S. Presidential race was a multi-media campaign. The candidates organized volunteers and engaged voters online, their partisans created video clips, and the campaigns themselves used numerous existing and new platforms to share the word and get out the

* Fellow, Berkman Center for Internet & Society at Harvard Law School and Silicon Flatirons Center at University of Colorado School of Law. Contact wendy@seltzer.org for the latest version. The author founded and leads the Chilling Effects Clearinghouse, <http://www.chillingeffects.org/>, from which much of the evidence is drawn. The author thanks participants in Harvard’s Berkman Center luncheon series, Silicon Flatirons and the University of Colorado Law School colloquium, and the IP Scholar Works-in-Progress for helpful discussion and comments on earlier drafts, and, particularly, her colleagues and collaborators on the Chilling Effects project.

© 2010 Wendy Seltzer. Reproduction permitted under the Creative Commons Attribution 3.0 License, <http://creativecommons.org/licenses/by/3.0/>

vote.¹ The McCain-Palin campaign reached out to voters by establishing a “channel” on the YouTube platform,² posting video clips where viewers could subscribe to the feed.

But in October, just weeks before the general election, several videos were removed from the McCain campaign’s YouTube channel, replaced with the terse advisory: “This video is no longer available due to a copyright claim.” The videos in question? Campaign advertisements. The claimants included CBS News, Fox News, the Christian Broadcasting Network, and NBC News, each apparently alleging that the ads infringed their copyrighted television programs.³ The McCain-Palin campaign wrote an impassioned letter to YouTube:

We write . . . to alert you to a problem that has already chilled this free and uninhibited discourse [O]verreaching copyright claims have resulted in the removal of non-infringing campaign videos from YouTube, thus silencing political speech [O]ur advertisements or web videos have been the subject of [Digital Millennium Copyright Act] takedown notices regarding uses that are clearly privileged under the fair use doctrine. The uses at issue have been the inclusion of fewer than ten seconds of footage from news broadcasts in campaign ads or videos, as a basis for commentary on the issues presented in the news reports, or on the reports themselves.⁴

McCain-Palin’s counsel urged YouTube to make an exception for the videos posted by political candidates and campaigns.⁵ He suggested that YouTube commit to a legal review of these political videos and decline to remove clearly non-infringing material, rather than taking

1. See LEE RAINIE & AARON SMITH, PEW RESEARCH CTR., THE INTERNET AND THE 2008 ELECTION (June 15, 2008), <http://www.pewinternet.org/Reports/2008/The-Internet-and-the-2008-Election.aspx> (reporting that 46% of Americans have used the Internet for political purposes, and 39% of online Americans have used the Internet to gain access to primary political documents and observe campaign events).

2. *McCain Hopes To Attract Young Voters*, CBS NEWS, May 12, 2008, <http://www.cbsnews.com/stories/2008/05/12/politics/main4087471.shtml>.

3. See Nate Anderson, *Fixing DMCA Takedown Problems Through Shaming, Legal Reform*, ARS TECHNICA, <http://arstechnica.com/old/content/2008/10/fixing-dmca-takedown-problems-through-shaming-legal-reform.ars> (last updated Oct. 20, 2008, 11:35 PM).

4. Letter from Trevor Potter, Gen. Counsel, McCain-Palin 2008, to Chad Hurley, CEO, YouTube, et al. (Oct. 13, 2008), available at <http://amlawdaily.typepad.com/amlawdaily/files/mccain-letter-20081013.pdf>.

5. *Id.* at 2 (“[W]e believe it would consume few resources — and provide enormous benefits — for YouTube to commit to a full legal review of all takedown notices on videos posted from accounts controlled by (at least) political candidates and campaigns.”)

down and insisting on a Digital Millennium Copyright Act (“DMCA”) waiting period of ten to fourteen business days.⁶

YouTube responded the next day, but said its hands were tied by the DMCA and that it would not play favorites among the many videos posted.⁷ YouTube’s counsel stated that “YouTube is merely an intermediary in this exchange, and does not have direct access to . . . critical information” regarding copyright ownership and infringement.⁸ McCain could counter-notify, sue, and use the court of public opinion to pillory the complainants, but he could not get the videos restored to YouTube before the expiration of the statutory delay — less than a month before November’s general election. Senator (or President) McCain could also, YouTube suggested, work to change the law so that others were not ensnared by it in the future.⁹

If there was ever a clear case of non-infringing fair use — speech protected by the First Amendment — this should have been it: a political candidate, seeking to engage in public multimedia debate, used video snippets from the television programs on which the issues were discussed.¹⁰ Following standard DMCA-induced policy,¹¹ however, YouTube never examined the legal validity of the underlying copyright complaint.¹² So long as the claimant sent notice to YouTube compliant with the statute’s formal requirements,¹³ YouTube would respond expeditiously to remove the claimed video. Why risk even the remote chance of litigation for someone else’s video?

As a result of unreviewed copyright complaints, political speech was removed from the McCain-Palin YouTube channel for more than a week at the height of election season, although “[i]t is well known that the public begins to concentrate on elections only in the weeks immediately before they are held.”¹⁴ Nor were the challenges limited to Republicans. The Obama-Biden campaign also lost access to You-

6. *Id.* at 2. *But cf.* Richard A. Posner, *Free Speech in an Economic Perspective*, 20 SUFFOLK U. L. REV. 1, 10 (1986) (“[T]here is no clear demarcation between political speech and other speech, once the purpose of protecting political speech is understood to be the preservation of political competition.”).

7. Letter from Zahava Levine, Chief Counsel, YouTube, to Trevor Potter, Gen. Counsel, McCain-Palin 2008, at 2–3 (Oct. 14, 2008), *available at* <http://wendy.seltzer.org/media/youtube-letter-20081014.pdf> (“We try to be careful not to favor one category of content on our site over others, and to treat all of our users fairly . . .”).

8. *Id.* at 2.

9. *Id.* at 3.

10. Under the fair-use factors, *see* 17 U.S.C. § 107 (2006), the McCain Nation channel used the clips for commentary in a non-commercial context, used a factual work, took only a small portion of a broadcast, and used the content in a manner that would not affect the broadcasters’ market.

11. *See DMCA Policy*, YOUTUBE, http://www.youtube.com/t/dmca_policy (last visited Dec. 21, 2010).

12. *See* Letter from YouTube to McCain-Palin 2008, *supra* note 7, at 2.

13. *See* 17 U.S.C. § 512(c)(3) (2006).

14. *Citizens United v. FEC*, 130 S. Ct. 876, 895 (2010).

Tube videos¹⁵ and the group Progress Illinois found its channel disabled.¹⁶

No court would have granted an injunction to suppress McCain's political speech. "The First Amendment does not permit laws that force speakers to retain [an] attorney . . . or seek declaratory rulings before discussing the most salient political issues of our day."¹⁷ Ultimately, therefore, the *Citizens United* Court struck down Congress's attempt to limit the impact of money on politics, holding that a statute banning corporate electioneering (funding speech shortly before an election) violated the First Amendment.¹⁸ As the Court observed, "First Amendment standards . . . 'must give the benefit of any doubt to protecting rather than stifling speech.'"¹⁹

The government, defending the campaign finance law at issue in *Citizens United*, argued that the law could be limited by FEC interpretation to avoid protected speech. The Court rejected the government's argument, holding that requiring speech to be pre-approved was equivalent to banning it. The Court's denunciation applies equally to censorship effected through the DMCA takedown scheme:

"[m]any persons, rather than undertake the considerable burden (and sometimes risk) of vindicating their rights through case-by-case litigation, will choose simply to abstain from protected speech — harming not only themselves but society as a whole, which is deprived of an uninhibited marketplace of ideas." Consequently, "the censor's determination may in practice be final."²⁰

Thus, rather than "prolong the substantial, nation-wide chilling effect" of the FEC's uncertain restrictions on speech, the Court invalidated the section.²¹

While the McCain-Palin campaign had other avenues than YouTube for spreading its message — including the campaign's own

15. See Steve McClellan, *YouTube Pulls Obama Spot*, ADWEEK, Oct. 1, 2008, http://www.adweek.com/aw/content_display/news/agency/e3i226441afd9c0206fb4262e8f1dec94f7.

16. See David Ardia, *Fox Television Forces Shutdown of Progress Illinois' YouTube Channel*, CITIZEN MEDIA LAW PROJECT (Jan. 6, 2009), <http://www.citimedialaw.org/blog/2009/fox-television-forces-shutdown-progress-illinois-youtube-channel>.

17. *Citizens United*, 130 S. Ct. at 889.

18. See *id.* at 913 (holding 2 U.S.C. § 441b facially unconstitutional for prohibiting corporations and unions from using their general treasury funds to make independent expenditures for speech that is an electioneering communication or express advocacy).

19. *Id.* at 891 (quoting *FEC v. Wis. Right to Life, Inc.*, 551 U.S. 449, 469 (2007) (Roberts, C.J.)).

20. *Id.* at 896 (alteration in original) (citations omitted) (quoting, respectively, *Virginia v. Hicks*, 539 U.S. 113, 119 (2003); *Freedman v. Maryland*, 380 U.S. 51, 58 (1965)).

21. *Id.* at 894.

website, with possibly stouter-hearted hosting service — most individuals do not have these alternatives. In the face of a DMCA takedown, they would “choose simply to abstain from protected speech.”²² For smaller, less powerful speakers, the initial takedown following a DMCA-backed copyright complaint strikes a final and fatal blow.

Federal law, through copyright and the DMCA, is responsible for this restriction on Internet speech. This is true even though the DMCA relies upon private enforcement, because of the incentive structure the DMCA creates for online intermediaries. As the Court observed in *Citizens United* and in earlier campaign finance cases, depriving speakers of opportunities for publication and dissemination by pressuring distribution points can be tantamount to banning speech: “Were the Court to uphold these [electioneering] restrictions, the Government could repress speech by silencing certain voices at any of the various points in the speech process.”²³

Writing for the Court in *Citizens United*, Justice Kennedy concluded that the election laws — which restricted the financing of speech, and thus the opportunity to speak — functioned “as the equivalent of prior restraint” on speech.²⁴

The same reasoning should apply to the barriers that copyright secondary liability and the DMCA pose to speakers. These barriers function as a prior restraint by inducing the necessary service provider²⁵ to take down speech before, and often in the absence of, a judicial determination of its infringing nature.

Each week, blog posts are redacted, videos deleted, and web pages removed from Internet search results based upon private claims of copyright infringement. The “safe harbor” provision of the DMCA encourages service providers to respond to copyright complaints with content takedowns, which assure the service providers immunity from liability while diminishing the rights of their subscribers and users. The law’s shield for service providers becomes, paradoxically, a sword against the public, which depends upon these providers as platforms for speech.

The DMCA provides limited legal process for an accused infringer. The law offers service providers protection from copyright liability if they remove material “expeditiously” in response to unveri-

22. *Cf. id.* at 896 (discussing chilling effects from FEC regulation).

23. *Id.* at 898 (citing *McConnell v. FEC*, 540 U.S. 93, 251 (2003) (Scalia, J., concurring in part and dissenting in part)).

24. *Id.* at 896.

25 As described in Part III, *infra*, the DMCA sets out several categories of “service provider,” including providers of hosting, search, and conduit services. *See* 17 U.S.C. § 512(k)(1) (2006) (defining “service provider” to mean a “provider of online services or network access, or the operator of facilities therefor,” and including conduit services). I use the generic term “service provider” to refer to any of these actors interchangeably unless a more precise reference is required.

fied complaints of infringement.²⁶ Even if the accused infringer responds with counter-notification asserting non-infringement, the DMCA requires the service provider seeking protection from liability to keep the material offline for more than a week.²⁷

If this takedown procedure took place through the courts, it would trigger First Amendment scrutiny as a prior restraint — silencing speech before an adjudication of unlawfulness. But because DMCA takedowns are privately administered through service providers, they have not received such constitutional scrutiny despite their high risk of error.

I add to prior scholarly analysis of the conflict between copyright and the First Amendment by showing how the copyright notice-and-takedown regime operates in the shadow of the law, silencing speech indirectly through private intermediaries where the government could not do so directly. In the wake of *Citizens United*, why can copyright law remove political videos from public reach when campaign finance law must not?

This Article argues for greater constitutional scrutiny. The DMCA's indirect chilling effect upon speech harms the public no less than if the government wrongly ordered the removal of lawful online material directly. Indeed, because sending a DMCA takedown notification costs copyright claimants less than filing a complaint in federal court and exposes claimants to few risks, it invites more frequent abuse and error than standard copyright adjudications. I describe several cases of error in detail. The indirect nature of the chill on speech should not shield the legal regime from challenge.

When non-infringing speech is taken down, not only does its poster lose an opportunity to reach an audience, the public loses the benefit of hearing that lawful speech in the marketplace of ideas.²⁸ Because of the DMCA's pressure, the poster's private incentives to counter-notify and the host's incentives to support challenged speech are often insufficient to support an optimal communication environment for the public. Instead, this set of incentives produces a blander but not significantly less copyright-infringing information space.

Copyright claimants assert that the expedited process of the DMCA is critical to suppress infringement in the highly-networked digital world. While many instances of infringement are properly targeted for takedown under the DMCA, I argue that the correctness of some takedowns does not excuse error elsewhere, nor the failure to undertake careful examination of the rate and costs of error. I therefore recommend rebalancing speech protection and copyright to reduce erroneous takedown.

26. 17 U.S.C. § 512(c)(1)(C) (2006).

27. *Id.* § 512(g).

28. *See Abrams v. United States*, 250 U.S. 616, 630 (1919).

Part II surveys the legal, economic, and structural sources of the DMCA's chilling effects on speech. Part III examines the First Amendment doctrines that should guide lawmaking, with a critique of copyright's place in speech law. Part IV reviews the history and mechanics of the DMCA and provides examples of chilled speech and a few instances of limited warming. Finally, Part V engages current policy debates and proposes reform to protect online speech better.

II. DMCA'S CHILL AS "PRIOR RESTRAINT BY PROXY"

The DMCA flips the defaults on speech. Ordinarily, speech remains available until someone files — and wins — a lawsuit or negotiates a settlement.²⁹ In contrast, a DMCA takedown forces a speaker to act to reassert the lawfulness of his speech through a counter-notification, or if he wants uninterrupted speech, a lawsuit.³⁰ This added cost operates as censorship. The poster who thinks his quotation is fair use may be willing to post but not to file a sworn counter-notification, just as the claimant for veterans' benefits who thinks he is engaged in loyal criticism may nonetheless object to having his benefits conditioned on a loyalty oath.³¹

The DMCA safe harbors may help the service provider and the copyright claimant, but they hurt the parties who were absent from the copyright bargaining table³²: the smaller individual and non-profit speakers using the Internet. The threat of secondary liability induces service providers to comply with the DMCA's notice-and-takedown provisions, making it more difficult for speakers to post material that challenges someone who can potentially make a claim to copyright.³³

Some of the examples of abuse cited below are extreme, but they are not isolated. The frequency of error and its bias against speech represents a structural problem with secondary liability and the

29. See Thomas I. Emerson, *The Doctrine of Prior Restraint*, 20 LAW & CONTEMP. PROBS. 648, 648 (1955) ("In constitutional terms, the doctrine of prior restraint holds that the First Amendment forbids the Federal Government to impose any system of prior restraint, with certain limited exceptions, in any area of expression that is within the boundaries of that Amendment.").

30. See *infra* Part III (describing the mechanics of the DMCA notice-and-takedown regime).

31. See *Speiser v. Randall*, 357 U.S. 513 (1958) (invalidating as a violation of the First Amendment a loyalty oath requirement on veterans claiming tax exemptions).

32. For the copyright law negotiation as a "bargaining table" from which the general public is excluded, see JESSICA LITMAN, *DIGITAL COPYRIGHT* 126 (2001). Of course, many of those asserting copyright claims via the DMCA are small or non-commercial entities. But even if claimants were evenly balanced across the size-wealth spectrum, the targets affected by takedowns are disproportionately the poorer speakers who also have fewer alternatives to the Internet to make their voices heard and less influence with their service providers to ward off claims.

33. Secondary liability, or intermediary liability, refers to holding one party (the service provider) liable for the acts of another (the poster/speaker). See *infra* Part III.

DMCA: the DMCA makes it too easy for inappropriate claims of copyright to produce takedown of speech. It encourages service providers to take down speech on notice even if the notice is factually questionable or flawed. It encourages copyright owners to use copyright claims as a route to expeditious takedown. The DMCA thus enhances the power of the claimant over the alleged infringer.

First, even good-faith uses of the DMCA are problematic when the underlying law is uncertain. Copyright law and its fair use provisions are far from bright-line. DMCA notices force service providers to confront fact-specific fair use disputes that even courts would be unable to decide on summary judgment. “The task [of fair use analysis] is not to be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis.”³⁴ Because the copyright doctrine is hazy, good-faith complainants may file erroneous DMCA claims against fair uses of their copyrights.

Compounding the problem, the promise of rapid takedown creates an incentive for copyright claimants to file dubious takedown claims. The mechanism is cheap for the claimant, more expensive for the respondent,³⁵ and if the process stops after the claim stage (as it often does) the complained-of material remains offline. And unless the complaint is so groundless that it can give rise to a lawsuit against the complainant,³⁶ a non-infringing poster has no legal or practical recourse against bogus claims.

Consequently, the DMCA is systematically susceptible to abusive claims. When the benefits of unlawful activity exceed the risk-

34. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994). The Supreme Court has ruled on fair use several times in the last two decades alone, on decisions that “were overturned at each level of review.” 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 13.05 n.10 (2010).

35. The respondent who counter-notifies must make “a statement under penalty of perjury that the subscriber has a good faith belief that the material was removed or disabled as a result of mistake or misidentification of the material to be removed or disabled,” and must consent to jurisdiction in United States courts. Moreover, the incremental cost of an additional takedown notice is low for the complainant who has already determined the legal form, prepared a template, and identified the DMCA agent for service, or hired a third-party service to send takedown notices on its behalf. *See Repeat Senders*, CHILLING EFFECTS, <http://www.chillingeffects.org/weather.cgi?NewsID=643> (last visited Dec. 21, 2010) (counting the most prolific senders of takedown notices). The targets of notice are more often confronting the process for the first time, since the law calls for policies of “termination” of repeat infringers. 17 U.S.C. § 512(i)(1)(A).

36. Under 17 U.S.C. § 512(f) (2006), the complainant can be sued and made to pay the poster’s costs and attorneys’ fees if shown to have “knowingly materially misrepresent[ed]” the infringing nature of material. Few cases to date have imposed these sanctions. *Compare Online Policy Group v. Diebold, Inc.*, 337 F. Supp. 2d 1195 (N.D. Cal. 2004) (granting plaintiff’s claim under 17 U.S.C. § 512(f)), *with Rossi v. Motion Picture Assoc. of Am.*, 391 F.3d 1000 (9th Cir. 2004) (dismissing plaintiff’s tortious interference claim relating to defendant’s filing of a DMCA takedown notice).

weighted penalties for being caught, parties can get away with much borderline or unlawful activity.³⁷

Furthermore, the intermediating role of service providers adds a layer of indirection to the operation of the law, and slows access to information that could help market forces correct errors. Several factors complicate individual efforts to protect speech: the divergence of provider incentives from those of any of the parties to the underlying copyright dispute, the parties' differing risk assessments, and the information and monitoring costs of a principal-agent relationship. Legal scholarship has only recently acknowledged the detriments to speech that intermediation and gatekeeper liability may cause. When intermediaries are necessary for speech, pressures on these private relations take on First Amendment significance.

A. Error Costs of the DMCA

The First Amendment requires courts and legislatures to anticipate error in application of the law, and to err on the side of allowing speech rather than restricting it.³⁸ As Frederick Schauer argues, the expected error and uncertainty in the legal process combine with this strong First Amendment preference to produce “definitional balances” that intentionally permit some unwanted activity, as in defamation, obscenity, and incitement law.³⁹ By favoring copyright claimants, the DMCA skews this balance in the wrong direction.

If a copyright-enforcement system worked perfectly, infringement would be detected and stopped rapidly without impairing the creation of and access to non-infringing works.⁴⁰ Real-world copyright enforcement can fall short of this ideal in two ways: by failing to stop infringement or by stopping non-infringing speech. Law aims to deal with the tradeoffs between these two types of error, false negatives and false positives.⁴¹

37. See Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994).

38. See, e.g., *New York Times Co. v. Sullivan*, 376 U.S. 254, 270–72 (1964) (observing that even factually incorrect statements deserve First Amendment protection lest fear of mistake stifle public debate).

39. Frederick Schauer, *Fear, Risk and the First Amendment: Unraveling the “Chilling Effect,”* 58 B.U. L. REV. 685, 686–87 (1978).

40. The argument regarding optimal enforcement is distinct from the choice between rules and standards, focusing instead on accurate enforcement of whatever rule or standard has been chosen. See generally Dan L. Burk, *Muddy Rules for Cyberspace*, 21 CARDOZO L. REV. 121 (1999).

41. See Robert G. Bone, *Enforcement Costs and Trademark Puzzles*, 90 VA. L. REV. 2099, 2123 (2004) (“Two different types of error must be considered separately: false positives and false negatives. Generally, a false positive occurs when a party obtains a result he should not have obtained and a false negative occurs when a party fails to obtain a result that he should have obtained.”); Alfred C. Yen, *Internet Service Provider Liability for Subscriber Copyright Infringement, Enterprise Liability, and the First Amendment*, 88 GEO. L.J. 1833, 1869–70 (2000). In constitutional jurisprudence, these errors are frequently referred to

Often, trying to eliminate one type of error will exacerbate the other.⁴² If law's overriding priority were to stop online infringement — no matter how many innocent, non-infringing users and uses were mistakenly punished — it could reduce the false negatives of uncaught infringers at the expense of catching more non-infringers: more dolphin caught in the fishing nets.⁴³ A law aimed at this goal might give copyright claimants a free hand to demand takedown of claimed infringements, force service providers to terminate the accounts of users accused of infringement, or hold service providers strictly liable for any infringement occurring on or through their networks. Such a law would exaggerate the false positives and increase the takedown or silencing of non-infringing speech. Strict liability for service providers, in particular, might cut online infringements, but it would substantially reduce non-infringing speech, particularly from individual or non-commercial speakers, as service providers would demand high fees or bonds to indemnify themselves against the possibility of infringement and liability.

If, on the other hand, law were concerned above all with safeguarding non-infringing use and publication of non-infringing works, it would focus on limiting the false positives, minimizing the lawful users caught by copyright complaints and non-infringing posts mistakenly removed. Such a law might immunize service providers from liability for their users' activity, putting the providers under no obligation to respond to copyright complaints based on user-generated content.⁴⁴ It might even convert copyright protection from the current property rule — backed by speech-removing injunction — into a liability rule.⁴⁵ Such a regime might offer claimants money damages and possible injunction only after a court found infringement, lest pre-

as "overbreadth" (or "overinclusion") and "underinclusion." See Kenneth W. Simons, *Overinclusion and Underinclusion: A New Model*, 36 UCLA L. Rev. 447, 510 (1989) (categorizing the due process challenges).

42. See Bone, *supra* note 41, at 2124. As Bone notes:

The expected cost of each type of error depends upon two factors: the frequency of the error and the social cost produced by that type of error. The reason to distinguish between the two different types of error is that they may produce different social costs. Many legal rules reduce the frequency of one type of error only to increase the frequency of the other.

Id. Both types of error could be reduced if the process were made more accurate, but not without the addition of resources. See Kaplow, *supra* note 37, at 308.

43. See Dennis Roddy, *The song remains the same*, PITTSBURGH POST-GAZETTE, Sept. 14, 2003, at B1 (quoting RIAA spokesperson Amy Weiss as saying that "[w]hen you fish with a net, you sometimes are going to catch a few dolphin"), available at <http://www.post-gazette.com/columnists/20030914edroddy0914p1.asp>.

44. *Cf.* 47 U.S.C. § 230 (2006) (immunizing service providers from liability for the non-intellectual property activity of their users).

45. See generally Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972) (describing liability and property rules as means of protecting entitlements).

liminary takedowns or preliminary injunctions disable access to lawfully-posted material. This would probably lead to more un-chilled posting at the cost of an increase in false negatives — infringing posts that were not quickly remedied.

The DMCA’s drafters steered clear of those extremes, but the law’s tendency toward erroneous removal predominates over its meager mechanisms for correcting these errors.

B. Intermediation

The DMCA exacerbates the problems of intermediation inherent to secondary liability, thus presenting a species of principal-agent problem.⁴⁶ Because the agent-service provider does not share all the benefits of the principal-poster, the agent lacks a similarly strong incentive to take risks in defending posted material in the face of a complaint. Incentives may be misaligned with social interests.⁴⁷

The service provider is a third-party intermediary on the critical path to online speech.⁴⁸ Service providers are imperfect agents for their poster-principals.⁴⁹ These intermediaries to online speech likely have different incentives and risk sensitivities from their users, and the additional layer they represent increases information costs. The DMCA plays upon these divergences to suppress speech and deprive the public of positive externalities from speech.

First, interests and incentives differ between poster and service provider. A principal-agent relationship poses challenges because

46. In this economic analysis, the principal is the one who wants something posted, while the agent is the one who acts on his behalf to accomplish it. See Lewis A. Kornhauser, *An Economic Analysis of the Choice Between Enterprise and Personal Liability for Accidents*, 70 CALIF. L. REV. 1345, 1346 (1982) (distinguishing economic from legal agency). This is a weak agency relationship, as the poster’s only lever of control is to pay hosting fees or provide eyeball-worthy content. Note that this is the obverse of the vicarious liability assessment, in which the service provider is implied-in-law to be a principal to its poster-agent if it derives financial benefit from and maintains the “right and ability to supervise” the posted material. Cf. *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913, 931 (2004); *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004 (9th Cir. 2001).

47. Cf. Michel van Eeten & Johannes M. Bauer, *Emerging Threats to Internet Security: Incentives, Externalities and Policy Implications*, 17 J. CONTINGENCIES & CRISIS MGMT., 221, 223 (2009) (suggesting that misalignment of private incentives among users, service providers, and software vendors produces too little investment in online security).

48. See Wendy Seltzer, Remarks, *The Politics of Internet Control & Delegated Censorship*, 102 AM. SOC’Y INT’L L. PROC. 45, 45 (2008).

49. In particular, the agency relationship is imperfect because many posters get no binding commitment from their service providers. Terms of service give the service provider the option to terminate the relationship at any time. See Sandra Braman & Stephanie Roberts, *Advantage ISP: Terms of Service as Media Law*, 5 NEW MEDIA & SOC’Y 422 (2003). Nonetheless, since the poster needs one or more service providers to be heard, he is stuck with these imperfect agents. See sources cited *infra* note 54 for why the market does not provide better alternatives. On the principal-agent problem, see generally PAUL ROBERT MILGROM & JOHN ROBERTS, *ECONOMICS, ORGANIZATION & MANAGEMENT* (1992); JOHN WINSOR PRATT & RICHARD ZECKHAUSER, *PRINCIPALS AND AGENTS* (1985).

each of the parties is motivated by his or her own self-interest, however broadly understood, and the principal-poster can only imperfectly direct the agent-service provider.⁵⁰ Naturally, the poster is more invested in his or her speech than is the host, for whom it is but one of many posts. Even if the host is charging the subscriber or reaping advertising revenue from pageviews, the two will at best be sharing the value created.⁵¹ Moreover, copyright law encourages providers to be skeptical of a party who is too willing to pay extra for guarantees — the “financial benefit” attributable to that specific activity might be deemed a trigger for vicarious liability,⁵² or advertisement of takedown-resistant services may be seen as inducing infringement.⁵³ In the current market, a provider might fear that offering takedown-resistant services would lead to adverse selection, concentrating in their subscriber base the knowing, intentional infringers since they would most anticipate needing such services and therefore be willing to pay.⁵⁴

Second, the poster-service-provider relationship is prone to information asymmetries: the poster is in a better position to know the copyright status of her work. While an automated scan may be able to identify a match between posted material and known copyright-claimed works,⁵⁵ it cannot determine the relevant *copyright* status of the posted work.⁵⁶ What appear at first to be wholesale infringements may in fact be postings authorized by the copyright owner,⁵⁷ fairly

50. Where the hosting fees are cheap or free, the poster-principal has little leverage apart from threatening to take his business elsewhere, which may even look attractive to the service provider if the poster appears prone to incur liabilities for the service provider.

51. Economists speak of the double marginalization problem — namely, that each party in the vertical production or distribution chain claims separate returns — as a prompt to vertical integration. From a speech perspective, however, we would hardly want to force all would-be speakers to become their own service providers.

52. *See* *Fonovisa, Inc. v. Cherry Auction, Inc.*, 76 F.3d 259, 261–64 (9th Cir. 1996) (finding vicarious liability based on the right and ability to control and financial benefit directly connected to the infringing activity). The DMCA imports the *Fonovisa* standard. *See* 17 U.S.C. § 512(c) (2006) (“A service provider shall not be liable for monetary relief . . . if the service provider . . . does not receive a financial benefit directly attributable to the infringing activity, in a case in which the service provider has the right and ability to control such activity . . .”).

53. *See* *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913 (2005).

54. *See* MILGROM & ROBERTS, *supra* note 49, at 149; George A. Akerloff, *The Market for “Lemons”: Qualitative Uncertainty and the Market Mechanism*, 84 Q. J. ECON. 488 (1970).

55. *See, e.g.,* *Content Management*, YOUTUBE, <http://www.youtube.com/t/contentid> (last visited Dec. 21, 2010).

56. *See* David Abrams, *More Chilling than the DMCA — Automated Takedowns, CHILLING EFFECTS* (March 17, 2010), <http://chillingeffects.org/weather.cgi?WeatherID=634> (describing audio fingerprinting and its use in automated content blocking).

57. *See* Memorandum of Law in Support of Defendants’ Motion for Summary Judgment at 39–52, *Viacom Inc. v. YouTube, Inc.*, 540 F. Supp. 2d 461 (S.D.N.Y. 2008) (No. 07 Civ. 2103), available at http://www.google.com/press/pdf/20100318_google_viacom_youtube_memorandum.pdf.

used excerpts,⁵⁸ or even originals from which the claimant's copy was derived.⁵⁹ The poster is in a better position to know the facts behind the posting, but will have difficulty convincing the service provider she is telling the truth. Moreover, the DMCA exacerbates the information problem by encouraging service providers not to look at their users' posted content in advance of a notification lest they acquire "actual knowledge" of infringement (or be sued on that claim).⁶⁰ A service provider whose every support call costs money is unlikely to investigate copyright ownership, authorization, or potential fair use defenses to an infringement claim rather than simply pulling the complained-of content or link.

Some scholars — and many in the entertainment publishing industry — argue that service providers should act as copyright gatekeepers, either because they deem the host to share some responsibility for the infringement or, more instrumentally, because hosts are positioned to stop infringing activity more rapidly or cheaply.⁶¹ These analyses focus on the property harms of copyright infringement, and they tend to minimize the public costs — in reduced speech and access — of intermediary enforcement.⁶² Yet Fred Yen, who succinctly defines the theory of "enterprise liability" as the view

58. See Lawrence Lessig, *Update on Warner Music (UPDATED) (AGAIN)*, LESSIG (Apr. 30 2009, 4:15 PM), http://www.lessig.org/blog/2009/04/update_on_warner_music.html; Mike Masnick, *Bogus Copyright Claim Silences Yet Another Larry Lessig YouTube Presentation*, TECHDIRT (Mar. 2, 2010, 4:26 AM), <http://www.techdirt.com/articles/20100302/0354498358.shtml>.

59. See *Disagreement over License for Scrapbook Designs*, CHILLING EFFECTS (Sept. 10, 2009), <https://www.chillingeffects.org/derivative/notice.cgi?NoticeID=28439>; *Graphic Designer Gets Cease and Desist from Former Client*, CHILLING EFFECTS (Dec. 16, 2009), <http://www.chillingeffects.org/copyright/notice.cgi?NoticeID=31515>.

60. The DMCA declares the safe harbor available if, inter alia, the service provider "does not have actual knowledge that the material or an activity using the material on the system or network is infringing." 17 U.S.C. § 512(c)(1)(A)(i) (2006).

61. Jonathan Zittrain describes the "gatekeeping" model, derived from Reinier Kraakman's work, without endorsing it. Jonathan Zittrain, *A History of Online Gatekeeping*, 19 HARV. J. LAW & TECH. 253, 256 (2006) ("Such liability asks intermediaries who provide some form of support to wrongdoing to withhold it, and penalizes them if they do not.") (citing Reinier H. Kraakman, *Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy*, 2 J.L. ECON. & ORG. 53, 53–54 (1986)). "These are intermediaries of various kinds — generally those who carry, host, or index others' content — whose natural business models and corresponding technology architectures have permitted regulators to conscript them to eliminate access to objectionable material or to identify wrongdoers in many instances." *Id.* at 253–54.

62. See Niva Elkin-Koren, *Making Technology Visible: Liability of Internet Service Providers for Peer-to-Peer Traffic*, 9 N.Y.U. J. LEGIS. & PUB. POL'Y 15, 25 (2006) ("By enabling high quality copying at negligible cost and facilitating mass distribution of copies at the click of a mouse, digital networks elevated piracy to gigantic proportions. The cost of enforcing copyrights increased immensely . . ."); Assaf Hamdani, *Who's Liable for Cyberwrongs?*, 87 CORNELL L. REV. 901 (2002); Ronald J. Mann & Seth R. Belzey, *The Promise of Internet. Intermediary Liability*, 47 WM. & MARY L. REV. 239, 250 (2005) ("[T]he time has come for the Internet to grow up and for Congress and the businesses that rely on the Internet to accept a mature scheme of regulation that limits the social costs of illegal Internet conduct in the most cost-effective manner.").

that “[e]nterprises that create risk should bear the burden of that risk as a cost of doing business,”⁶³ concludes that such liability should not be applied to service providers because of First Amendment considerations: “a broad application of enterprise liability may be deeply problematic because enterprise liability easily becomes liability without limit.”⁶⁴ Instead, “proper interpretation of copyright law leaves plenty of weapons available against the individuals who commit clear copyright infringement without dragging [service providers] into the fray.”⁶⁵

C. The DMCA and the Economics of Speech

To the service provider, the DMCA offers the choice between streamlined self-censorship on the one hand, and, on the other, case-by-case determination of the liability risks and costs of defending against a claim. In the ordinary course, risk aversion prevails, especially when the stake is another party’s speech. To the public, potentially valuable speech is lost in the shuffle.

Why does the market fail to correct for this error? Much online speech is non-commercial,⁶⁶ and its hosting is free or low-margin, without room for insurance.⁶⁷ The costs of potential copyright liabil-

63. Alfred C. Yen, *Internet Service Provider Liability for Subscriber Copyright Infringement, Enterprise Liability, and the First Amendment*, 88 GEO. L.J. 1833, 1856 (2000) (“Such cost internalization is more than just fair. It encourages risk creators to take precautions against loss, it provides compensation for victims, and it spreads the costs among all who benefit from the risk-creating activity.”).

64. *Id.* at 1856.

65. *Id.* at 1893.

66. See AMANDA LENHART ET AL., PEW INTERNET & AM. LIFE PROJECT, SOCIAL MEDIA AND YOUNG ADULTS 2, 20, 23 (2010), http://www.pewinternet.org/~media/Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplevels.pdf (as of September 2009, 73% of online American teens used online social network websites; 86% of teen social network users post comments to friends’ pages); AMANDA LENHART ET AL., PEW INTERNET & AM. LIFE PROJECT, CONTENT CREATION ONLINE (2004), http://www.pewinternet.org/~media/Files/Reports/2004/PIP_Content_Creation_Report.pdf (finding, in a 2003 survey, that “44% of adult Internet users have used the Internet to publish their thoughts, respond to others, post pictures, share files and otherwise contribute to the explosion of content available online. 21% of Internet users say they have posted photographs to Web sites. 13% of Internet users maintain their own Web sites”); SACHA WUNSCH-VINCENT & GRAHAM VICKERY, DIRECTORATE FOR SCI., TECH. & INDUS., OECD, PARTICIPATIVE WEB AND USER-CREATED CONTENT: WEB 2.0, WIKIS AND SOCIAL NETWORKING (2007), <http://www.oecd.org/dataoecd/57/14/38393115.pdf>.

67. Many popular web hosting service are free to the end-user. See Jean-Samuel Beuscart & Kevin Mellet, *Business Models of the Web 2.0: Advertising or the Tale of Two Stories* COMM. & STRATEGIES (SPECIAL ISSUE), November 2008, at 165, available at <http://ssrn.com/abstract=1374448> (characterizing these services as multi-sided platforms, often selling user traffic to advertisers.). According to Alexa, six of the ten websites with the largest audiences are Web 2.0 sites that permit users to post content at no charge: Facebook.com (2), Youtube.com (3), Blogger.com (7), Wikipedia.com (8), QQ.com (9), and Twitter.com (10). See *Top Sites*, ALEXA, <http://www.alexa.com/topsites> (last visited Dec. 21, 2010).

ity — statutory damages of \$750 to \$150,000 per work infringed or higher actual damages⁶⁸ — might be more than a service provider is comfortable letting a customer indemnify it against, even if the customer were inclined to purchase “takedown-proof” hosting to support a tolerance for risk higher than the host’s.

Moreover, the public benefit from access to speech is an externality or “spillover” whose value is generally not captured by the speaker, nor, therefore, by the hosting costs a speaker is willing to pay.⁶⁹ Criticisms and parodies benefit the public, providing value both to their direct readers and to those whose democratic society is shaped for the better as a result of such dialogue.⁷⁰ The public at large has no good way to pay into the system to support these speech-derived benefits,⁷¹ but we could subsidize such benefits broadly by diminishing the risks and costs of speech.⁷²

Prior to the DMCA, each participant in the chain could at least make independent decisions about copyright compliance and liability risk. The poster might post because she was confident of her legal right or self-censor because of uncertainty or fear of liability; the service provider might make both initial entry decisions (whether to join this market) and subsequent decisions (whether to learn about customer behavior, take down or ignore upon notice); and the copyright claimant might evaluate the copyright claims against antagonists, costs of asserting claims, and risks of wrongful assertion. The poster was, of course, dependent on one or more service providers to allow him to post, link to, or access content, unless he was large enough to

68. 17 U.S.C. § 504(c) (2006) (permitting the copyright owner to seek statutory damages of \$750 to \$30,000 per work infringed, increasing up to \$150,000 in cases of willful infringement).

69. See Brett M. Frischmann, *Speech, Spillovers, and the First Amendment*, 2008 U. CHI. LEGAL F. 301, 301–02; Brett M. Frischmann & Mark A. Lemley, *Spillovers*, 107 COLUM. L. REV. 257, 258 (2007). Rather than decrying uncaptured external benefits, Frischmann and Lemley celebrate technological spillovers for leaving social value for others to innovate upon. The same is true in the cultural space, where we would not want to see an economic accounting precede every conversation (and stop many).

I am not contending that all speech is valuable. Hate speech and defamation have negative externalities. See Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for the Information Age*, 91 B.U. L. REV. (forthcoming July 2011). On balance, as the First Amendment recognizes, an open speech environment produces more public benefit than the alternative.

70. See Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 4–5 (2004); Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 356–58 (1999); Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 288–89, 385–86 (1996).

71. Advertisements, micropayments, and tip jars notwithstanding, most of us do not pay for a great deal of the online speech we read and watch.

72. See MILGROM & ROBERTS, *supra* note 49, at 145 (describing the public goods problem); Frischmann, *supra* note 69, at 301 (describing subsidies, taxes, and direct payments among the options for compensating externalities).

operate his own backbone peering connection. Concerns of copyright secondary liability were only beginning to hit the mainstream.

Post-DMCA, we see the parties' decision-making processes realigned. While the poster faces a similar choice at posting time, the service provider sees a new set of decision points. At startup, the effect may be positive: the potential service provider may be emboldened by the opportunity of the safe-harbor provisions to offer services that would seem too risky if not for the clearly delineated procedures for avoiding liability.⁷³ Thus, the new service provider is likely to register a DMCA agent and institute the policies required by § 512(i). On an ongoing basis, however, this preparation invites § 512(c)(3) takedown notices, which put the service provider on notice and compel the "expeditious response" described in § 512(c)(1)(C). Building the system invites its use. The DMCA does not force service providers to avail themselves of its harbor, but shapes their risk assessment so that almost all do, even in cases where, objectively, no harbor appears necessary.

The safe-harbor and takedown regime, moreover, is not evenhanded. It distorts the speech environment by excessively removing challenged speech.⁷⁴ On balance, this set of incentives produces a blander information space without reducing infringement. The so-called pirates, interested in sharing popular mass-media, will always be able to exploit darknet economies — with so many motivated mice, a few will always remain out of the cat's reach.⁷⁵ The posters of non-mass content, by contrast, will be stymied, tripped up by administrative costs and barred from reposting by "repeat infringers" provi-

73. Fred von Lohmann argues that the DMCA has had a positive impact in this regard, pointing to the explosion of user-generated content sites launched post-1998, such as YouTube, Flickr, Blogger, and Vimeo. Von Lohmann argues that sites were able to obtain clearance to launch and attract investment because the DMCA gave them a straightforward way to assert their lawfulness without pre-screening every post against a potentially infinite pool of sources it might infringe. Fred von Lohmann, Senior Staff Attorney, Elec. Frontier Found., Lecture at the University of Colorado Silicon Flatirons: Digital Copyright and Innovation Online: A Little Dose of Optimism (Oct. 13, 2009). We lack the benefit of a controlled experiment, however. If the DMCA's safe-harbor opportunities had not been available at the same point in time, similar sites might have launched nonetheless, asserting they bore no liability as mere carriers of user-posted content. Google Book Search, for example, was launched on generic claims of fair use, but is currently being reshaped by a class action lawsuit and settlement proposal. See James Grimmelmann, *D Is for Digitize: An Introduction*, 55 N.Y.L. SCH. L. REV. 11 (2010).

74. This ability to remove content with mere notice makes the DMCA like a forbidden "heckler's veto," whereby anyone who dislikes speech can make it more costly to host. Cf. *Reno v. ACLU*, 521 U.S. 844, 880 (1997) (invalidating the Communications Decency Act, because, among other reasons, the requirement not to communicate indecent speech to "specific persons" "would confer broad powers of censorship, in the form of a 'heckler's veto,' upon any opponent of indecent speech").

75. See Fred von Lohmann, *Measuring the DMCA Against the Darknet: Implications for the Regulation of Technological Protection Measures*, 24 LOY. L.A. ENT. L. REV. 635, 636–40 (2004).

sions.⁷⁶ This means that copies of *The Dark Knight* will spread more easily than transformative commentary on it, and *Saturday Night Live* skits will be more widely available than parodies (or political advertisements) that build upon them.⁷⁷ The consequence is a vicious circle, whereby the continued presence of infringing materials increases demand for harsher enforcement, which further increases the costs of hosting challenged material, yet fails to stop the infringement. The tax of DMCA takedowns distorts the speech environment, biasing it against a particular kind of “troublesome” speech.

III. FIRST AMENDMENT PROBLEMS

A. Copyright and the First Amendment

Describing the “paradox” of copyright’s speech regulation in 1970, Professor Melville Nimmer concluded that the conflict between copyright and the First Amendment was more apparent than real.⁷⁸ Instead, copyright’s “definitional balance,” the idea-expression dichotomy, provided sufficient breathing room for free expression.⁷⁹ Speakers were properly outside of copyright when they appropriated others’ *ideas*, while they could make few First Amendment-relevant claims to merit direct copying of others’ *expressions*.⁸⁰ Following Nimmer’s lead, courts have regularly used the idea-expression dichotomy and copyright’s fair use exceptions to explain away First Amendment concerns.⁸¹

Harper & Row v. Nation Enterprises turned on an unusual set of facts: *Time Magazine* had purchased the first publication rights to excerpts from former President Gerald Ford’s memoirs and was preparing for its quote-filled article to appear a week before the book’s

76. See 17 U.S.C. § 512(i) (2006) (requiring that a service provider have “adopted and reasonably implemented . . . a policy that provides for the termination in appropriate circumstances of . . . repeat infringers” for the safe harbor to apply). Copyright challenges may exaggerate the speech inequalities described by Neil Weinstock Netanel. See Neil Weinstock Netanel, *Market Hierarchy and Copyright in Our System of Free Expression*, 53 VAND. L. REV. 1879, 1899 (2000) (“Copyright fosters speech hierarchy.”). The popular get more popular, while the marginal are marginalized further.

77. More technically, we might say that the elasticity of supply of mass-interest works is less than that of niche works, so that original ideas are more likely to be squeezed out by higher risks and costs. Cf. Richard A. Posner, *Free Speech in an Economic Perspective*, 20 SUFFOLK U. L. REV. 1, 20 (1986) (suggesting that speech taxes or tax-equivalent burdens would drive out the “marginal producer of ideas — as the producer of a new idea will often be”).

78. Melville B. Nimmer, *Does Copyright Abridge the First Amendment Guarantees of Free Speech and Press?*, 17 UCLA L. REV. 1180, 1180–81 (1970).

79. *Id.* at 1190.

80. *Id.*

81. See *Eldred v. Ashcroft*, 537 U.S. 186, 215 (2003); *Harper & Row v. Nation Enters.*, 471 U.S. 539, 560 (1985); *Zacchini v. Scripps-Howard Broad. Co.*, 433 U.S. 562, 577 (1977).

publication when *The Nation*, operating from a purloined manuscript, scooped it. *The Nation*'s 2250-word article quoted 300 to 400 words from the memoir in which Ford described Nixon's resignation and his pardon. *Time* canceled its article and refused to pay the balance it owed to Harper & Row.⁸²

Sued for copyright infringement, *The Nation* argued that its use was fair news reporting: "not only the facts contained in Mr. Ford's memoirs, but 'the precise manner in which [he] expressed himself [were] as newsworthy as what he had to say.'"⁸³

The Court did not agree. Finding that a "public concern" exception could swallow copyright for public figures' accounts of noteworthy events, the Court fell back upon the bulwark of the idea-expression dichotomy:

In view of the First Amendment protections already embodied in the Copyright Act's distinction between copyrightable expression and uncopyrightable facts and ideas, and the latitude for scholarship and comment traditionally afforded by fair use, we see no warrant for expanding the doctrine of fair use to create what amounts to a public figure exception to copyright.⁸⁴

Harper & Row did not warrant so broad a dictate. The First Amendment issues were not paramount in the case. *Time* and *Harper & Row* were both planning publication, so *The Nation*'s action advanced release of the "news" of Ford's description of events by only a few weeks. The case does not reveal any urgent political debate to which the early release contributed, nor any time-sensitive commentary that *The Nation* could provide only with early quotes. At the same time, *The Nation*'s publication effectively substituted for *Time*'s, sapping the value from the first publication rights Ford's publisher had sold.

Thus, presented with what seemed a minor impingement upon speech but a blow to copyright, the Court acted to preserve copyright. But while the Court's response might have fit the particular circumstances of this case, its broad terms, and its characterization of copyright as "the engine of free expression"⁸⁵ led future courts to see copyright as practically immune from First Amendment scrutiny.

Yet even Nimmer was not certain that proper definitions of idea and expression could balance away all expressive interest in copy-

82. *Harper & Row*, 471 U.S. at 539.

83. *Id.* at 556.

84. *Id.* at 560.

85. *Id.* at 589.

ing,⁸⁶ and since 1970, the countervailing interests have grown. As copyright's expansion and extension put increasing pressure on the fulcrum, scholars have returned to the question and found more substantial conflict.⁸⁷ The Court has not yet done so, holding in *Eldred v. Ashcroft* that the definitional balance and the safeguard of fair use protections still sufficed to save term extensions from First Amendment overreach. "[W]hen . . . Congress has not altered the traditional contours of copyright protection, further First Amendment scrutiny is unnecessary."⁸⁸

This debate has tended to focus on the law as properly applied: whether a careful interpretation of the idea-expression dichotomy and proper application of the fair use doctrine prevents copyright from encroaching on the domain of free speech. This is not the only question that must be asked, however. Often, the law is interpreted without this due care, sometimes by individuals censoring themselves to avoid crossing the line, sometimes by claimants mistaken or overeager about their rights, sometimes by courts, and frequently by service providers responding to takedown notices. As I describe in Part IV, *infra*, these errors are all too frequent in the context of online takedown demands. Because copyright's subject matter is speech, the effect of copyright errors silencing protected speech is of constitutional concern.⁸⁹

86. Nimmer felt that for news photographs, leeway for the copying of ideas alone did not satisfy the First Amendment balance: "No amount of words describing the 'idea' of the [My Lai] massacre could substitute for the public insight gained through the photographs It would be intolerable if the public's comprehension of the full meaning of My Lai could be censored by the copyright owner of the photographs." 4-19E Nimmer on Copyright § 19E.03. In today's multimedia environment, Nimmer's nugget of concern takes on greater importance. When news is made in televised speeches, one who wants to comment or criticize needs the immediacy of the footage. Online, one may often link to materials, but also want to save copies in case the originally linked version changes. One may need the original, as in the *Diebold* case discussed below, for its proof of authenticity, not its creative expression.

87. See Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 3 (2004); Mark A. Lemley & Eugene Volokh, *Freedom of Speech and Injunctions in Intellectual Property Cases*, 48 DUKE L.J. 147, 150–51 (1998); Neil Weinstock Netanel, *Locating Copyright Within the First Amendment Skein*, 54 STAN. L. REV. 1, 2–4, 85–86 (2001); Rebecca Tushnet, *Copy this Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 YALE L.J. 535, 538, 587–90 (2004). Courts' inattention to First Amendment issues is "unfortunate, because most intellectual property rules — copyright law, trademark law, right of publicity law, and trade secret law — are speech restrictions: They keep people from publishing, producing, and performing the speech that they want to publish, produce, and perform." Eugene Volokh, *Freedom of Speech and Intellectual Property: Some Thoughts After Eldred*, 44 Liqueurmart, and Bartnicki, 40 HOUS. L. REV. 697, 698 (2003).

88. *Eldred v. Ashcroft*, 537 U.S. 186, 215, 221 (2003).

89. There is some speech on the other side of the balance as well, on the argument that copyright's "engine of free expression" depends on the ability to stop infringement, so erroneous underblocking is also speech-impairing. I suggest that this interest is more attenuated, particularly in cases of quotation and derivative work as distinct from wholesale appropriation.

Recent scholarship, drawing upon the wealth of expression fostered by the combination of cheap, powerful multimedia creation and fast connectivity, has pushed the First Amendment envelope even further.⁹⁰ Rebecca Tushnet argues for “copying as free speech,” arguing that fair use “transformation” is not necessary to make copying socially valuable.⁹¹

B. Prior Restraints on Speech

The takedowns resulting from DMCA notifications bear many of the hallmarks of prior restraints on speech⁹²: they are imposed to limit speech before any adjudication on the merits of the copyright claims. While takedowns are effected by private actors, service providers are acting “in the shadow of the law,”⁹³ motivated by the state action that established copyright liability and the DMCA. Government cannot insulate itself from responsibility for this abridgment of free speech by routing its influence through third-party service providers.

“Any system of prior restraints of expression comes to [the] Court bearing a heavy presumption against its constitutional validity.”⁹⁴ “If it can be said that a threat of criminal or civil sanctions after publication ‘chills’ speech, prior restraint ‘freezes’ it at least for the time.”⁹⁵ The prior restraint doctrine’s greatest utility is as a bright line, keeping questions of administrative pre-clearance of speech off the table rather than entertaining case-by-case judgments of the restraints’ utility.⁹⁶ Thus, a libelous or obscene publication may not be enjoined before publication, and only after heightened scrutiny may its publisher be made to pay after-the-fact damages. As Mark Lemley and Eugene Volokh have noted, it is already difficult to square the presumption of

90. See YOCHAI BENKLER, *THE WEALTH OF NETWORKS* (2006); JAMES BOYLE, *THE PUBLIC DOMAIN* (2008); LAWRENCE LESSIG, *REMIX* (2008) (advancing a view of “Read-Write” culture); Julie E. Cohen, *The Place of the User in Copyright Law*, 74 *FORDHAM L. REV.* 347, 370 (2005) (describing the “situated user [who] appropriates cultural goods . . . [for] consumption, communication, self-development, and creative play”); William W. Fisher, *Property and Contract on the Internet*, 73 *CHI. KENT L. REV.* 1203, 1217–18 (1998) (describing “semiotic democracy”); Jennifer Rothman, *Liberating Copyright: Thinking Beyond Free Speech*, 95 *CORNELL L. REV.* 463 (2010) (identifying a “liberty interest” in copying beyond First Amendment analysis).

91. Tushnet, *supra* note 87, at 540, 562.

92. See, e.g., *Near v. Minnesota*, 283 U.S. 697, 736, 738 (1931) (invalidating a statute that provided for injunction of a “malicious, scandalous and defamatory” periodical as an unconstitutional restraint on publication).

93. Cf. Robert H. Mnookin & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 *YALE L.J.* 950 (1979).

94. *Bantam Books, Inc. v. Sullivan*, 372 U.S. 58, 70 (1963).

95. *Neb. Press Ass’n v. Stuart*, 427 U.S. 539, 559 (1976).

96. See Emerson, *supra* note 29, at 648 (“[T]he doctrine of prior restraint is, in some important respects, more precise in its application than most of the other concepts that have developed out of the First Amendment.”).

“irreparable harm” and frequent issuance of preliminary injunctions in copyright cases with this doctrine.⁹⁷

In his early and influential review of prior restraint doctrine, Thomas Emerson outlined the key characteristics distinguishing prior restraint from subsequent punishment: (1) breadth, (2) timing and delay, (3) propensity toward adverse decision, (4) limited procedure, (5) limited opportunity for public appraisal and criticism, (6) the “dynamics of prior restraint,” (7) certainty and risk, and (8) effectiveness.⁹⁸ These elements contribute to prior restraint’s particular threat to free expression. The DMCA notice-and-takedown regime exhibits similar flaws: (1) Overbreadth: facially conformant but erroneous notices routinely prompt takedown; any posted content is potentially susceptible. (2) Delay: the ten-to-fourteen-business-day takedown can be timed strategically, to remove speech at the time of greatest impact to an ongoing debate. (3) Nearly all general-purpose providers take down content almost automatically upon receipt of a conformant notice. (4) The poster generally receives no notice or opportunity to respond until after content is taken down, and may receive few specifics even then; the only opportunity to contest is through counter-notice, which is biased against the poster, or in court. (5) Private actions are even less open to public appraisal than those of government censors; the indirect nature of the regulation diverts criticism.⁹⁹ (6) The posting of information regarding DMCA agents and procedures invites their use.¹⁰⁰ (7) The risk involved with filing a counter-notification is made to appear greater than the risk of initial posting. (8) On a case-by-case basis, the takedown scheme is effective. Almost every instance targeted by a notification is removed, and yet, in gross, the system fails to stop infringement of mass content because more targets re-appear from new sources.¹⁰¹

Prior restraint doctrine thus inclines us toward procedural safeguards for speech as a curb on administrative censorial discretion, as a motive to get more speech to the “marketplace of ideas,” and as a pro-

97. Lemley & Volokh, *supra* note 87, at 150. The limitation on preliminary injunctions in patent cases, announced in *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006), has only begun to erode the presumption of injunctive relief in copyright cases.

98. Emerson, *supra* note 29, at 656–59.

99. See LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 95–98 (2000).

100. Cf. John C. Jeffries, Jr., *Rethinking Prior Restraint*, 92 YALE L.J. 409, 422 (1982) (“[A]dministrative preclearance requires a bureaucracy of censorship . . . [T]here are powerful institutional pressures to justify one’s job, and ultimately one’s own importance, by exaggerating the evils which suppression seeks to avoid.”).

101. See von Lohmann, *supra* note 75. For one recent example, a Google search for “Avatar movie download” returns pointers to Fox DMCA takedowns listing more than 3,000 distinct URLs. See *Fox DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Jan. 4, 2010), <http://www.chillingeffects.org/notice.cgi?sID=17619>. It nonetheless remains possible to find the film for unauthorized download.

tection, even if incomplete, from the chill.¹⁰² In particular, it focuses us on what Jeffries terms the “institutional structure” of speech law,¹⁰³ or what Lessig calls “architecture.”¹⁰⁴ Monaghan identifies in the Supreme Court case law a “First Amendment due process,” requiring judicial determination before speech is restrained.¹⁰⁵ “[F]irst [A]mendment rights are fragile and can be destroyed by insensitive procedures; in order to completely fulfill the promise of those cases, courts must thoroughly evaluate every aspect of the procedural system which protects those rights.”¹⁰⁶

Some commentators have expressed skepticism about the prior restraint doctrine’s cohesion, or prior restraint’s distinction from subsequent punishment.¹⁰⁷ Some argue that subsequent punishment is a more severe restriction on speech than prior restraint, because the would-be speaker loses his opportunity to test his speech before facing criminal sanctions.¹⁰⁸ If an injunction is less painful than a jail term, then a speaker might prefer that a court enjoin him. The First Amendment, therefore, ought to be as concerned with the self-censorship that subsequent punishment induces as it is with the censorship of prior restraint.

This Article need not choose between these positions because the DMCA procedure provides the worst of both. To the poster, the service provider’s summary takedown looks like an injunction, but without even the benefit of judicial review. At least a temporary restraining order or preliminary injunction requires a court hearing; DMCA takedowns might get only the review of an overworked paralegal. Unless the poster intervenes with a trip to court, expressive ma-

102. See Henry P. Monaghan, *First Amendment “Due Process,”* 83 HARV. L. REV. 518, 519 (1969) (“Like the substantive rules themselves, insensitive procedures can ‘chill’ the right of free expression.”); Martin H. Redish, *The Proper Role of the Prior Restraint Doctrine in First Amendment Theory*, 70 VA. L. REV. 53, 58 (1984).

103. See Jeffries, *supra* note 100, at 422.

104. Lawrence Lessig, Address at [www9: Cyberspace’s Architectural Constitution](http://cyber.law.harvard.edu/works/lessig/www9.pdf) (June 12, 2010), available at <http://cyber.law.harvard.edu/works/lessig/www9.pdf>.

105. See Monaghan, *supra* note 102, at 520–22.

106. *Id.* at 551.

107. See, e.g., OWEN M. FISS, *LIBERALISM DIVIDED: FREEDOM OF SPEECH AND THE MANY USES OF STATE POWER*, 136 (1996) (“The prior restraint doctrine . . . should not be seen as a full or coherent expression of free speech values, but rather as a strategic device capable of effectuating a compromise, the chief value of which is negative — to block a decision against speech.”); Stephen R. Barnett, *The Puzzle of Prior Restraints*, 29 STAN. L. REV. 539, 558–60 (1977) (challenging the presumption that prior restraint is more chilling than threat of subsequent punishment); Jeffries, *supra* note 100, at 425 (focusing attention on “the structure of [the law’s] administration”); William T. Mayton, *Toward a Theory of First Amendment Process: Injunctions of Speech, Subsequent Punishment, and the Costs of the Prior Restraint Doctrine*, 67 CORNELL L. REV. 245, 245–46 (1982); Martin Scordato, *Distinction Without a Difference: A Reappraisal of the Doctrine of Prior Restraint*, 68 N.C. L. REV. 1 (1989).

108. See Ariel L. Bendor, *Prior Restraint, Incommensurability, and the Constitutionalism of Means*, 68 FORDHAM L. REV. 289 (1999); Steven Alan Childress, *The Empty Concept of Self-Censorship*, 70 TUL. L. REV. 1969 (1996); Redish, *supra* note 102, at 71.

terial will be offline or inaccessible for a minimum of ten business days even under the best of counter-notification circumstances. Moreover, takedown fails to ensure that the poster will not be sued for infringement, because the protective aspects of the safe harbor apply only to the service provider.

Many of these problems arise from the interaction of law and the incentives of the actors. While the DMCA does not force service providers to avail themselves of its harbor, it does shape their risk assessment. As a result, almost all service providers take advantage of the safe harbor, even in cases where objectively no harbor appears necessary. The DMCA is at least partially to blame for this inclination on the part of service providers to take advantage of the safe harbor.¹⁰⁹

Even if the DMCA and secondary copyright liability cannot be invalidated as a classic prior restraint, many of the reasons for disfavoring prior restraints apply here as well. The DMCA deprives the public of both access to speech that would ultimately be ruled lawful and the judicial certainty that would come from earlier adjudication of many of these disputes. As an interim solution, the public might even prefer a brief, less formal adjudication.¹¹⁰

C. Understanding Chilling Effects

First Amendment law accounts for the likelihood of error in the law's assessment or application through a "chilling effects" analysis.¹¹¹ The law prohibits states from imposing liability without fault for defamation not because it favors falsehoods, but because it embodies the concern that a stricter rule would inhibit truthful reporting. Even when the core of a law's prohibition is unquestionably unprotected speech, the chilling effects on protected speech around its edges may give significant reason for invalidating or reining it in further.

As Frederick Schauer explains, the "chilling effect" doctrine derives from recognition that the legal process is uncertain and that the First Amendment expresses a preference for errors in favor of speech rather than those that restrict it.¹¹² "The doctrine flows from the relationship between our recognition of the inevitability of error and our preference for a particular type of error; and it is the existence of this

109. The threat of secondary liability underlying the DMCA's effect on the behavior of service providers is not poised to change. The current Court seems unlikely to invalidate secondary liability on First Amendment grounds, given that it has approved more direct governmental pressure on speech, such as government-funded viewpoint control, see *Rust v. Sullivan*, 500 U.S. 173 (1991), and the government-funded censorship of library Internet access, see *United States v. American Library Ass'n*, 539 U.S. 194 (2003).

110. Cf. Mark A. Lemley & R. Anthony Reese, *A Quick and Inexpensive System for Resolving Peer-to-Peer Copyright Disputes*, 29 CARDOZO ARTS & ENT. L.J. 1 (2005) (proposing informal adjudication in the context of peer-to-peer infringement cases).

111. See Schauer, *supra* note 39.

112. Schauer, *supra* note 39, at 689.

relationship . . . [that] justifies the formulation of substantive rules in this area.”¹¹³

Typically, the chilling effect doctrine is concerned with excessive promotion of self-censorship. An individual may refrain from speech that the law does not intend to target because of fear that the law will adversely affect him. He may do so for several reasons — first, because he fears that he will be found liable though he has done no wrong; second, because he anticipates the cost of defense will be high, even if he trusts the result will ultimately be correct; and third, because he doubts the absolute correctness of his position and he faces high costs if he is found to be incorrect. The law’s chilling effect on an individual is thus a function of the likelihood of erroneous enforcement, the costs of litigation, and the magnitude of the harm from punishment.

The chilling effect doctrine has formed part of the Court’s analysis against many speech-affecting laws, particularly those challenged for vagueness, overbreadth, and improper burden-shifting. The Court first used the term “chilling effect” in the 1965 case *Dombrowski v. Pfister*, where the concern for “chill” was invoked against an overbroad “subversive activities” law:

Because of the sensitive nature of constitutionally protected expression, we have not required that all of those subject to overbroad regulations risk prosecution to test their rights. For free expression — of transcendent value to all society, and not merely to those exercising their rights — might be the loser. . . . By permitting determination of the invalidity of these statutes without regard to the permissibility of some regulation on the facts of particular cases, we have, in effect, avoided making vindication of freedom of expression await the outcome of protracted litigation. Moreover, we have not thought that the improbability of successful prosecution makes the case different. The chilling effect upon the exercise of First Amendment rights may derive from the fact of the prosecution, unaffected by the prospects of its success or failure.¹¹⁴

Schauer traces the doctrine’s origins back to the 1958 case *Speiser v. Randall*,¹¹⁵ which threw out a requirement that veterans take a loyalty oath in order to receive benefits because it unfairly shifted the burden

¹¹³ *Id.*

¹¹⁴ 380 U.S. 479, 486–87 (1965) (internal citations omitted).

¹¹⁵ 357 U.S. 513 (1958).

to speakers to justify the lawfulness of their speech.¹¹⁶ The doctrine recognizes that one cannot be assured that her rights will be vindicated cheaply and quickly by accurate courts:

There is always in litigation a margin of error, representing error in factfinding, which both parties must take into account. . . . Where the transcendent value of speech is involved, due process certainly requires in the circumstances of this case that the State bear the burden of persuasion to show that the appellants engaged in criminal speech.

The vice of the present procedure is that, where particular speech falls close to the line separating the lawful and the unlawful, the possibility of mistaken factfinding — inherent in all litigation — will create the danger that the legitimate utterance will be penalized. The man who knows that he must bring forth proof and persuade another of the lawfulness of his conduct necessarily must steer far wider of the unlawful zone than if the State must bear these burdens. This is especially to be feared when the complexity of the proofs and the generality of the standards applied provide but shifting sands on which the litigant must maintain his position. How can a claimant whose declaration is rejected possibly sustain the burden of proving the negative of these complex factual elements? In practical operation, therefore, this procedural device must necessarily produce a result which the State could not command directly. It can only result in a deterrence of speech which the Constitution makes free.¹¹⁷

The Court picks up the concern with both errors and burden-shifting in *New York Times v. Sullivan* by rejecting a libel standard in which the defendant must prove truth:

A rule compelling the critic of official conduct to guarantee the truth of all his factual assertions — and to do so on pain of libel judgments virtually unlimited in amount — leads to a . . . “self-censorship.” Allowance of the defense of truth, with the burden of

116. See Schauer, *supra* note 39 at 701.

117. 357 U.S. 513, 525–26 (1958) (internal citations omitted).

proving it on the defendant, does not mean that only false speech will be deterred.¹¹⁸

In defamation cases, therefore, courts accept the risk of unpunished falsehood. “[E]rroneous statement is inevitable in free debate, and . . . it must be protected if the freedoms of expression are to have the ‘breathing space’ that they ‘need . . . to survive.’”¹¹⁹ The false positive of lawful (truthful) speech deterred is worse than the false negative of some erroneous statements remaining unpunished.

Beyond its specific concern with error costs, the Court’s “chilling effect” analysis recognizes that the architecture of law can be blamed for its effect on parties beyond the courtroom and the four corners of a statute. The First Amendment places strict limits on those effects, whether direct or incidental. Laurence Tribe compares modern law to the “observer effect” in quantum physics.¹²⁰ The presence of background legal doctrine, like the presence of the observer of a quantum particle, changes activity. In cases such as *Sullivan v. New York Times*, *NAACP v. Claiborne Hardware Co.*, and *Hustler Magazine v. Falwell*:

[T]he Supreme Court held that first amendment principles were violated not by some state official’s act of censorship but by the overall shape of the state’s body of judge-made rules for awarding damages to people allegedly injured by speeches or publications. . . . [T]he Supreme Court’s entire development of the “chilling effect” doctrine over the past several decades . . . reflects a judicial recognition that widespread private behavior, in the form of self-censorship, can be directly traceable not only to particular enforcement actions by specific state officials

118. 376 U.S. 254, 279 (1964) (internal citations omitted). The *Sullivan* Court went on to note:

The fear of damage awards under a rule such as that invoked by the Alabama courts here may be markedly more inhibiting than the fear of prosecution under a criminal statute. . . . Whether or not a newspaper can survive a succession of such judgments, the pall of fear and timidity imposed upon those who would give voice to public criticism is an atmosphere in which the First Amendment freedoms cannot survive.

Id. at 277–78.

119. *Id.* at 271–72 (internal citations omitted, second omission in original); see also *Gertz v. Robert Welch*, 418 U.S. 323, 341 (1974) (“The First Amendment requires that we protect some falsehood in order to protect speech that matters.”).

120. Laurence H. Tribe, *The Curvature of Constitutional Space: What Lawyers Can Learn from Modern Physics*, 103 HARV. L. REV. 1 (1989).

but to the very existence of a set of rules or lines that the state stands ready to enforce or to draw.¹²¹

The intermediation of the service provider, by providing another subject of potential chill, adds to the First Amendment impact of the DMCA.

D. Chill, Intermediated

As Seth Kreimer says, “The [Supreme] Court was well aware that the coercive effect of indirect sanctions is magnified when deployed against intermediaries who transmit the work of others.”¹²² Kreimer traces modern First Amendment jurisprudence concerning intermediaries to the McCarthy Era, during which the Court reacted against indirect blacklisting.¹²³ When the House Un-American Activities Committee publicized lists of “communist sympathizers,” it expected and encouraged private actors to fire or shun the “Reds.”¹²⁴ In the late 1950s and 1960s, the Court held the government accountable for these indirect effects:

[T]he Court rejected the proposition that the First Amendment constrained only official efforts to criminally punish protected speech and association. Against the backdrop of the indirect sanctions of the McCarthy era, the Court recognized the potentially drastic effects of indirect gambits directed to vulnerable pressure points, and declared that First Amendment freedoms “are protected not only against heavy-handed frontal attack, but also from being stifled by more subtle governmental interference.”¹²⁵

Based on its experience with McCarthyism’s “subtle governmental interference,” the Supreme Court focused its attention on self-censorship, the pressures on intermediaries, and the potential inability of those harmed to bring suit.¹²⁶ In response, the Court developed protective doctrines, including proscriptions against overbroad legislation for its “chilling effect” on protected speech and pockets of immunity for truthful expression.¹²⁷ The Court broadened its view of First

121. *Id.* at 26.

122. Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 50 (2006).

123. *Id.* at 44.

124. *See id.* at 44–46.

125. *Id.* at 47 (quoting *Bates v. City of Little Rock*, 361 U.S. 516, 523 (1960)).

126. *Id.*

127. *Id.* at 47–48.

Amendment standing in order to enable hardier litigants such as associations to step in for weaker individuals.¹²⁸ Those safeguards are necessary today against commercially-motivated or malicious interference with expression.

Kreimer follows this thread of judicial concern with *intermediated* government action through defamation, obscenity, and broadcast indecency cases.¹²⁹ Throughout these cases, the Court raises the concern that because an intermediary bears the costs of investigating complaints, and its benefit from an individual item may be slight, it will be more averse to risk than the end-producer, viewer, or reader.¹³⁰ Thus, the defamation rule of *New York Times v. Sullivan* protects publishers, in order to give room to their speakers.¹³¹ By easing some of publishers' investigative burden and cost-based concerns, the "actual malice" standard enables them to make their presses available to "host" advertisers' speech, such as the "Heed Their Rising Voices" solicitation for Martin Luther King Jr.'s civil rights defense at issue in *New York Times*.¹³²

Again in obscenity law, the Court saw intermediaries making arguments on behalf of authors who needed publishers and bookstores through which to speak. "The *Bantam Books* Court observed that orders directed to intermediary distributors had the effect of suppressing the books of publishers who depended on those intermediaries to convey their books to the public"¹³³ and structurally deterring individuals from challenging the law: "The distributor who is prevented from selling a few titles is not likely to sustain sufficient economic injury to induce him to seek judicial vindication of his rights."¹³⁴

This thread continues through recent cases regarding election speech and financing. In *McConnell v. FEC*, Justice Scalia focuses most directly on intermediation and the role of money in securing intermediaries' services necessary to speech:

128. *Id.* at 51. Kreimer went on to note:

The McCarthy era warped the political culture of the United States by raising the risks of political action. . . . In response to this experience, a cluster of First Amendment doctrines evolved to bar legal mechanisms particularly likely to deter the less committed from political speech and association. These doctrinal structures established safe harbors in which unheroic citizens could still feel free to participate in discourse, to associate, and to facilitate the discourse of others.

Id. at 79–80.

129. Kreimer, *supra* note 122.

130. *Id.* at 50–55.

131. *Id.* at 54 (discussing *New York Times v. Sullivan*, 376 U.S. 254 (1964)).

132. See *Sullivan*, 376 U.S. at 256–57 (1964). The advertisement itself is available from the National Archives, at <http://arcweb.archives.gov/arc/action/ExternalDOSearch?searchExpression=2641477>.

133. Kreimer, *supra* note 122, at 53.

134. *Bantam Books v. Sullivan*, 372 U.S. 58, 64 n.6 (1963) (internal citations omitted) (permitting a publisher to challenge a state Morality Commission's advisory to bookstores that that some of the publisher's books were objectionable).

In any economy operated on even the most rudimentary principles of division of labor, effective public communication requires the speaker to make use of the services of others. An author may write a novel, but he will seldom publish and distribute it himself. A freelance reporter may write a story, but he will rarely edit, print, and deliver it to subscribers. To a government bent on suppressing speech . . . this mode of organization presents opportunities: Control any cog in the machine, and you can halt the whole apparatus. License printers, and it matters little whether authors are still free to write. Restrict the sale of books, and it matters little who prints them. Predictably, repressive regimes have exploited these principles by attacking all levels of the production and dissemination of ideas.¹³⁵

Transposed to the Internet, this same concern applies with great force to “Internet sources, such as blogs and social networking Web sites, [that] will provide citizens with significant information about political candidates and issues”¹³⁶ as well as education and entertainment.

Kreimer concludes that Internet “censorship by proxy” is similar to — and equally troubling as — this earlier pressure on intermediaries:

Analysis of efforts by the government to target weak links in Internet chains of communication thus takes place against the background of the long-standing position, rooted in the lessons of the McCarthy era, that “subtle interferences” and efforts to dissuade transmission by intermediaries constitute cognizable dangers to free expression, no less than threats of direct prosecution of speakers or listeners. The fact that these efforts enlist the cooperation of private parties makes them more, rather than less, dangerous in comparison to direct regulation. Private discretion is often less visible and less procedurally regular than public sanction.¹³⁷

135. 540 U.S. 93, 251 (2003) (Scalia, J., concurring in part, dissenting in part). While Scalia was in dissent in this part of *McConnell*, that position and much of its reasoning have now been adopted by the majority in *Citizens United v. FEC*, 130 S. Ct. 876, 913 (overruling *McConnell*).

136. *Citizens United*, 130 S. Ct. at 913.

137. Kreimer, *supra* note 122, at 65.

This concern naturally echoes that animating the prior restraint doctrine.¹³⁸ The opacity, procedural irregularity, and indirection of regulation through delegated private censors¹³⁹ parallel, if not exceed, those of the administrative censor.

Vulnerable intermediaries leave the speech that depends on them in a precarious state. Weaving together the threads of prior restraint, chilling effect doctrine, and intermediation can guide us to a legal safety net to rescue online speakers by limiting the chilling impact of secondary liability.¹⁴⁰ At the moment, that impact is serious, as is shown through a deeper investigation of the law and its application in practice.

IV. THE CHILL WINDS OF COPYRIGHT AND DMCA

In the early 1990s, as the Internet was officially opened to commercial activity, government was discussing it as the “National Information Infrastructure,” (“NII”) a name that belied the government’s limited understanding of the phenomenon.¹⁴¹ As Jessica Litman describes, the intellectual property working group assembled under Patent Commissioner Bruce Lehman felt that commerce needed some prodding, and that only expansive copyright grants and protections to authors would push retailers online.¹⁴² The group’s white paper concluded that “the full potential of the NII will not be realized if the education, information and entertainment products protected by intellectual property laws are not protected effectively when disseminated via the NII.”¹⁴³ Its view was less nuanced than Jane Ginsburg’s suggestion that greater control would enable new businesses.¹⁴⁴

138. See generally Emerson, *supra* note 29 (discussing the doctrine of prior restraint).

139. See Seltzer, *supra* note 48, at 45.

140. See *infra*, Part V.

141. The term “National Information Infrastructure” “encompass[ed] digital, interactive services now available, such as the Internet, as well as those contemplated for the future.” BRUCE A. LEHMAN ET AL., INTELLECTUAL PROPERTY AND THE NATIONAL INFORMATION INFRASTRUCTURE: THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS 2 n.5 (1995) [hereinafter NII WHITE PAPER]. For a fuller history of the early framing of the copyright question and its effect on the development of law, see LITMAN, *supra* note 32, at 89–110.

142. See LITMAN, *supra* note 32, at 93–94.

143. NII WHITE PAPER, *supra* note 141, at 10.

We see the same strategy transposed to the next new technology, digital television broadcasting, in debates over the proposed “broadcast flag.” Motion picture companies assert that high-definition digital television will take off only if their “high-value content” is available, and that their high-value content will be available only if it is granted additional protections against copying. Government regulators bought into this argument without fully considering that bowing to today’s commercial interests may well stifle even more promising commercial interests of tomorrow. See Wendy Seltzer, *The Broadcast Flag: It’s Not Just TV*, 57 Fed. Commun. L.J. 209 (2005).

144. See Jane C. Ginsburg, *From Having Copies to Experiencing Works: the Development of an Access Right in U.S. Copyright Law*, 50 J. OF THE COPYRIGHT SOC’Y OF THE

The group's white paper identified multiple barriers to commerce in the existing structures of the Net and copyright law. Its interpretation of copyright law made every Internet experience one of multiple copyright "reproductions," which were infringing if not otherwise licensed.¹⁴⁵ Further, according to the white paper, service providers faced potential liability for any of those infringements.¹⁴⁶

Because of the threat of liability, those in the nascent business of providing Internet service could be brought to the bargaining table to agree to "safe harbors" from these omnipresent liabilities — and to see the safe harbors offered by the DMCA's negotiators as valuable concessions in their favor.

The actual state of affairs was more complex. Service providers were assured neither immunity nor liability in the pre-DMCA world; they could have litigated to fix the bounds and obtain greater certainty. The DMCA was not the only approach to mitigating the risks associated with hosting third-party content.

The DMCA, as ultimately passed, does not incorporate the white paper's apocalyptic view of the necessary scope of copyright.¹⁴⁷ The DMCA does not change the underlying copyright liability of service providers who choose not to avail themselves of its safe harbors, or who try but fail to meet the safe-harbor conditions.¹⁴⁸

The On-Line Copyright Infringement Liability Limitation Act, codified at Section 512 of the DMCA, defines service providers of several types and sets conditions through which they can avoid secondary copyright liability. The safe harbor has four bays, one each for providers of connectivity, caching, hosting, and "information location" or linking services. Each bay has its own specifics and procedural prerequisites, but the overall structure is similar: service

USA 113, 113–14 (2003). Ginsburg acknowledges that the DMCA's changes to copyright create new rights. She argues those rights will promote new and valuable efforts by copyright creators, while refusing to expand copyright would diminish copyright's incentive. *See id.* at 122–23.

145. NII WHITE PAPER, *supra* note 141, at 64–66. Litman argues that this reading was an unwarranted extrapolation from a few questionable decisions, notably *MAI Systems Corp. v. Peak Computer*, 991 F.2d 511 (9th Cir. 1993). LITMAN, *supra* note 32, at 91–95; *see also* Pamela Samuelson, *The Copyright Grab*, WIRED, Jan 1996, <http://www.wired.com/wired/archive/4.01/white.paper.html>.

146. NII WHITE PAPER, *supra* note 141, at 114–124; *see also* Samuelson, *supra* note 145 ("The white paper asserts that every online service provider is already liable for all copyright infringement committed by its users, regardless of whether the service has reason to know about the infringement or takes reasonable steps to ensure that it won't occur.").

147. *See Ellison v. Robertson*, 357 F.3d 1072, 1076 (9th Cir. 2004) (recounting legislative history of the DMCA).

148. *See* 17 U.S.C. § 512(l) (2006) ("The failure of a service provider's conduct to qualify for limitation of liability under this section shall not bear adversely upon the consideration of a defense by the service provider that the service provider's conduct is not infringing under this title or any other defense."); *Ellison v. Robertson*, 357 F.3d at 1077 ("Claims against service providers for direct, contributory, or vicarious copyright infringement, therefore, are generally evaluated just as they would be in the non-online world.").

providers who comply with the specified conditions obtain immunity from liability for users' copyright infringement.¹⁴⁹ Service providers are defined in § 512(k)(1):

(A) As used in subsection (a), the term “service provider” means an entity offering the transmission, routing, or providing of connections for digital online communications, between or among points specified by a user, of material of the user's choosing, without modification to the content of the material as sent or received.

(B) As used in this section, other than subsection (a), the term “service provider” means a provider of online services or network access, or the operator of facilities therefor, and includes an entity described in subparagraph (A).¹⁵⁰

The additional conditions and protections applicable to providers in these categories vary from relatively straightforward (the connectivity provider is sheltered for automatic, user-requested transmission of material it does not modify)¹⁵¹ to Byzantine (it is unclear whether the caching provision has been of use to anyone).¹⁵²

Both the hosting and information-location bays of the safe harbor impose an additional requirement that “upon notification of claimed infringement,” the provider “respond[] expeditiously to remove, or disable access to, the material that is claimed to be infringing or to be the subject of infringing activity.”¹⁵³

A provider of hosting services wishing to keep its safe harbor options open must designate an agent to receive notifications of claimed infringement, make contact information available on its website, and register the agent with the Copyright Office.¹⁵⁴ Having done that, the service provider is protected by the safe harbor provided that it does not have actual knowledge of the infringement, does not benefit financially from infringing activity that it has the right and ability to control, and responds expeditiously to notifications of claimed infringement that follow the statutory form of § 512(c)(3).¹⁵⁵ Providers

149. See, e.g., 17 U.S.C. § 512(c) (2006). All providers are told they must have “a policy that provides for the termination in appropriate circumstances of subscribers and account holders of the service provider's system or network who are repeat infringers.” *Id.* § 512(i)(1)(A).

150. *Id.* § 512(k)(1).

151. *Id.* § 512(a) (“Transitory digital network communications.”).

152. *Id.* § 512(b) (“System caching.”).

153. *Id.* § 512(c)(1)(C).

154. *Id.* § 512(c)(2).

155. *Id.* § 512(c)(1).

who take down material pursuant to a copyright notification are immunized from liability.¹⁵⁶

The DMCA provides for counter-notification and replacement of erroneously removed material,¹⁵⁷ but these provisions do not parallel those of the initial notification. A provider can, consistent with its safe harbor immunity, replace material upon receipt of a counter-notification only after a period of ten to fourteen business days.¹⁵⁸ The subscriber who files a counter-notification must make “[a] statement under penalty of perjury that the subscriber has a good faith belief that the material was removed or disabled as a result of mistake or misidentification of the material to be removed or disabled.”¹⁵⁹ It is unclear whether an assertion that the material is authorized to be posted, or is fair use of another’s copyrighted material, satisfies the “mistake or misidentification” provision.¹⁶⁰ Further, the counter-notifier must agree to U.S. jurisdiction,¹⁶¹ a potential problem for non-U.S. parties. The service provider who accepts counter-notification avoids liability to its subscriber as well as to the copyright claimant.¹⁶² In practice, most service providers have placed clauses in their terms of service to preemptively avoid liability to their subscribers, making their restoration of material in compliance with the counter-notification provisions wholly optional.¹⁶³

Finally, the DMCA provides a remedy for one who is harmed by another who “knowingly materially misrepresents . . . that material or activity is infringing.”¹⁶⁴

As noted, the service provider is free at any time not to take shelter in the safe harbor. Rejecting the safe harbor does not create new liability or increase penalties, but leaves the service provider where it would have been under preexisting copyright law.¹⁶⁵

The DMCA was a product of its time. This became clear in the RIAA’s litigation against Verizon, begun in 2002, just a few years

156. *Id.* § 512(g)(1).

157. *Id.* § 512(g).

158. *Id.* § 512(g)(2)(C).

159. *Id.* § 512(g)(3)(C). By contrast, the only statement under penalty of perjury in the initial notification is the assertion “that the complaining party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed.” *Id.* § 512(c)(3)(A)(vi).

160. See Jennifer M. Urban & Laura Quilter, *Efficient Process or “Chilling Effects”?* *Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 621, 629–31 (2006). Compare *Lenz v. Universal Music Corp.*, No. C 07-3783 JF, 2010 U.S. Dist. LEXIS 16899 (N.D. Cal. Feb. 25, 2010) (finding mistake) with *Rossi v. Motion Picture Ass’n of Am. Inc.*, 391 F.3d 1000, 1004–05 (9th Cir. 2004) (finding no knowing misrepresentation).

161. 17 U.S.C. § 512(g)(3)(D) (2006).

162. *Id.* § 512(g)(1).

163. See Urban & Quilter, *supra* note 160, at 629.

164. 17 U.S.C. § 512(f).

165. See Urban & Quilter, *supra* note 160, at 629–31.

after the DMCA became law.¹⁶⁶ The RIAA invoked the subpoena provisions of § 512(h) of the DMCA in an attempt to obtain the names of Verizon subscribers alleged to have infringed record label copyrights via peer-to-peer filesharing.¹⁶⁷ The D.C. Circuit rejected the subpoenas on statutory grounds, saying § 512(h) applied only to parties who were providing hosting, location, or caching services, not to “mere conduits,” as Verizon was in relation to users of peer-to-peer networking.¹⁶⁸ “P2P software was ‘not even a glimmer in anyone’s eye when the DMCA was enacted,’” the court noted.¹⁶⁹

A. Errors and Pressures

In the ideal operation of the DMCA, a copyright holder who discovers her work has been infringed online sends a notice to the agent that the service provider has registered with the Copyright Office. Pursuant to § 512(c)(3)(A), the notice includes: identification of the work claimed to be infringed; identification of the material claimed to be infringing, with “information reasonably sufficient to permit the service provider to locate the material,”; contact information and signature of the complaining party; “a statement that the complaining party has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law,”; and “a statement that the information in the notification is accurate, and under penalty of perjury, that the complaining party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed.”¹⁷⁰

Imagine that a website, BookzNtextz.info, which is hosted by Mega-Service-Provider, is disseminating a novel without the author’s permission. In response, the author sends a letter to the registered agent of Mega-Service-Provider. She identifies her novel, Alice’s Easily Infringed Masterpiece, and gives the URL to her own website where the book can be purchased and the URL to the claimed infringement, <http://www.bookzntextz.info/AlicesEasilyInfringed/>. Alice includes the talismanic phrases “the complaining party has a good

166. See *Recording Indus. Ass’n of Am., Inc. v. Verizon Internet Servs., Inc.*, 351 F.3d 1229, 1238 (D.C. Cir. 2003) (“Congress had no reason to foresee the application of § 512(h) to P2P file sharing, nor did they draft the DMCA broadly enough to reach the new technology when it came along.”); *In re Verizon Internet Servs., Inc.*, 240 F. Supp. 2d 24, 29 (D.D.C. 2003).

167. *Recording Indus. Ass’n of Am., Inc.*, 351 F.3d at 1231.

168. *Id.* at 1236.

169. *Id.* at 1238 (quoting *Verizon Internet*, 240 F. Supp. 2d at 38).

170. 17 U.S.C. § 512(c)(3)(A)(i)–(vi) (2006). This copyright notification regime is far from ideal. The only statement made under penalty of perjury is the assertion of authority to act on behalf of an exclusive right holder. The notification leaves unsworn all allegations of infringement and identification. The notification is made to the service provider, not to a court, and it is not even required to be served on the alleged infringer.

faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law,” and attesting “that the information in the notification is accurate, and under penalty of perjury, that the complaining party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed.”¹⁷¹ Finally, Alice adds her contact information and signs the whole thing digitally or electronically.

To gain the protection of the DMCA safe harbor, Mega-Service-Provider, upon receiving the notice, expeditiously removes or disables access to the identified material.¹⁷² Mega-Service-Provider also “takes reasonable steps promptly to notify the subscriber that it has removed or disabled access to the material.”¹⁷³ If the subscriber counter notifies,¹⁷⁴ the service provider replaces the material or reinstates access ten to fourteen business days from receipt of the counter-notification.¹⁷⁵

For the case of Alice’s Easily Infringed Masterpiece, this rapid process works relatively well. It is lightweight: Alice can plug the results of her web searches into a form letter and obtain quick takedown of infringing copies. There are many spots where the process may deviate from the ideal, however. Alice may run her bots a bit too quickly,¹⁷⁶ sending notifications for Andy’s book review of the Easily Infringed Masterpiece; she may be sloppy in completing the DMCA form, listing a root URL rather than the location of a specific infringing work (a problem if a copy is posted to the forums of Bob’s Literary Criticism Realm, but the notification knocks out Bob’s whole site); she may be wrong about her legal rights, sending a claim about Carol’s Critical Reviews because he used her title (not copyrighted) or quoted passages from her opening chapter (fair use); finally, she may be deliberately malicious, sending a takedown regarding Diane’s Better Novel, perhaps after posting an excerpt from her text into the comment forum on that website.

Copyright claimants asserted unfounded claims before the DMCA as well, but the DMCA makes asserting such claims easier and the speech consequences more severe. This is so because service providers’ incentives do not match those of their users. Thus, while the service providers who receive takedown notifications *could* review each one closely with legal counsel (pushing the bounds of “expeditious”)

171. *Id.* § 512(c)(3)(A)(v)–(vi).

172. Pursuant to *id.* § 512 (c)(1)(A)(iii).

173. *Id.* § 512(g)(2)(A).

174. Pursuant to *id.* § 512(g)(3).

175. Pursuant to *id.* § 512(g)(2)(C).

176. Blake Reid’s analysis of recent complaints to Google by the International Federation of the Phonographic Industry (“IFPI”) suggests that if the agency were scanning twenty-four hours a day, seven days a week, it would be reviewing twenty URLs an hour — a pace likely kept only by bots (analysis on file with author).

and verify the content at each claimed location in comparison with the claimed original, it is cheaper and easier just to take down each and every claimed infringement. They might notify alleged infringers promptly and respond to counter-notifications, or they might skip those extra steps, instead insulating themselves from subscribers' suits through their terms of service.¹⁷⁷ More likely, given the margins in the industry, they would do the latter for all but their highest-paying customers.¹⁷⁸

Recall that if Bob wants to counter-notify, he must meet more stringent terms than did Alice for her initial notification. He must swear under penalty of perjury to his "good faith belief that the material was removed or disabled as a result of mistake or misidentification of the material to be removed or disabled"; give an address where he consents to the jurisdiction of Federal District Court, or, if he is outside the United States, consent to jurisdiction in any judicial district in which the service provider may be found; and say he will accept service of process from the person who made the initial notification or that person's agent.¹⁷⁹

Some of Alice's irritants may be hosted outside the United States or with uncooperative service providers. Still short of filing a lawsuit, she need not stop with one notice. She can send notifications to the service providers upstream of the offending hosts, the providers of aggregate hosting services or network connectivity to the sites hosting the allegedly infringing content. She can also notify search engines that their "information location tools" are referring users to allegedly infringing materials.¹⁸⁰ Any of these additional links in the chain between website and viewer may be more willing to act to remove material, disable access to it, or remove hyperlinks. Alternatively, the mere threat of an upstream attack may be enough to move the previously uncooperative service provider to cooperate.¹⁸¹

It is true that Alice could have made many of the same mistakes pre-DMCA, and the service providers could have been just as careful or careless in response to claims of infringement. But the DMCA

177. See Urban & Quilter, *supra* note 160 at 629.

178. See Beuscart & Mellet, *supra* note 67 (describing Web 2.0 service providers whose "technical and financial barriers to entry are low" and network externalities are strong, such that competition is to reach and maintain a critical mass of users).

179. 17 U.S.C. § 512(g)(3) (2006).

180. *Id.* § 512(d).

181. The threat of moving complaints to an "upstream" host can lend force to the ultimatum. Thus the Online Guitar Archive, OLGA.net, got a cease-and-desist letter from the National Music Publishers Association in July 2006 accompanied by a threatened DMCA notice: "Unless you remove all infringing material from your site voluntarily and within ten (10) days from the date of this notice, we will send you a notice, like that enclosed, in your capacity as an Internet Service Provider in accordance with the provisions of the Digital Millennium Copyright Act." In response, the site's operators pulled down all its guitar tablature files. See THE ON-LINE GUITAR ARCHIVE, <http://www.olga.net/> (last visited Dec. 21, 2010).

changes the calculus of both parties even without changing the underlying rules for liability in its absence. For the claimant, it makes the first step easier. Since the DMCA requires service providers who want possible immunity to register agents and post conspicuous contact information, the process of contact-gathering is simpler — many service providers make this information far easier to find than their customer service contacts, for example.¹⁸² Publicity surrounding the DMCA has advertised the usefulness of copyright claims for rapid takedown — often advertising it to those most likely to misuse it. For the provider, the DMCA's safe harbor offers a means of reducing risk.

The DMCA's choices present themselves to Alice, the claimant, well before filing a lawsuit. She need not hire a lawyer, pay a filing fee, or prepare for discovery. This situation makes it easier for her to stop actual infringement but also to err. Further, as we will see, the DMCA increases the error rate, not just the overall number of errors.¹⁸³

It has become popular to talk about Internet and online service providers as gatekeepers who can be enlisted in an orderly scheme of law enforcement online.¹⁸⁴ Although the Internet multiplies the number of speakers and speaking opportunities, and their opportunities for lawlessness, Internet architecture funnels that speech through relatively few hosts and information conduits. While it might be efficient to stop unlawful speech by cutting it off at the level of these hosts and conduits, it is impossible to do so without stopping a large amount of lawful speech.

Along with the extra incentives for copyright claimants come new pressures on service providers. To see why the gatekeepers overreact, we have to look at the situation from a service provider's viewpoint. DMCA claimants send most of their takedown notifications to providers of hosting for “information residing on systems or networks at direction of users”¹⁸⁵ and providers of “information location tools,”¹⁸⁶ commonly interpreted to mean search engines. The hosting category includes the small web host who runs a computing facility and allows users to create websites; the provider of blog hosting; the social networking site that allows users to create profiles and upload media; bulletin boards and online fora; group news sites; and other similar entities. The DMCA gives them a set of procedures to follow to posi-

182. Compare GETHUMAN, <http://gethuman.com/> (last visited Dec. 21, 2010) (relating accounts of the challenge of reaching customer service), with *Directory of Service Provider Agents for Notification of Claims of Infringement*, U.S. COPYRIGHT OFFICE, http://www.copyright.gov/onlinesp/list/a_agents.html (last visited Dec. 21, 2010).

183. In other words, the DMCA's error is not just an activity-level problem.

184. See Zittrain, *supra* note 61; DIRECTORATE FOR SCI., TECH. & INDUS., OECD, EXPERTS WORKSHOP ON INTERNET INTERMEDIARIES (2010), http://www.oecd.org/document/62/0,3343,en_2649_34223_44949886_1_1_1_37441,00.html.

185. 17 U.S.C. § 512(c) (2006).

186. *Id.* § 512(d).

tion themselves to respond to Alice's notification: They must register with the Copyright Office an agent to receive notifications of claimed infringement and make this agent's contact information accessible on their website.¹⁸⁷ They must further "[have] adopted and reasonably implemented, and inform[] subscribers and account holders of the service provider's system or network of, a policy that provides for the termination in appropriate circumstances of subscribers and account holders of the service provider's system or network who are repeat infringers."¹⁸⁸ They are also bound not to interfere with "standard technical measures,"¹⁸⁹ but this provision has not yet proved meaningful.

In return, if the provider responds expeditiously to notifications of claimed infringement that follow the statutory form of § 512(c)(3), does not have actual knowledge of the infringement, and does not benefit financially from infringing activity it has the right and ability to control, it is immunized from liability on both ends — from the copyright claimant for any part in the alleged infringement, and from the poster for claimed wrongful takedown.

An obvious explanation for the high rate of takedown would present itself if service providers were likely to face liability if they ventured outside of the safe harbor. The prior law does not fully support that proposition, however. The DMCA is clear, further, that "[t]he failure of a service provider's conduct to qualify for limitation of liability under this section shall not bear adversely upon the consideration of a defense by the service provider that the service provider's conduct is not infringing under this title or any other defense."¹⁹⁰ As the legislative history explains, Congress left underlying principles of liability unchanged for a provider who opted out from, or failed to comply with, the safe harbor.¹⁹¹ Under that prior law, liability was not a given; service providers had been held not to be direct or vicarious infringers based on their services' automatic copying of user-supplied material.¹⁹² These providers' possible contributory liability was

187. *Id.* § 512(c)(2).

188. *Id.* § 512(i)(1)(A).

189. *Id.* § 512(i)(1)(B).

190. *Id.* § 512(l).

191. *See, e.g.*, S. REP. NO. 105–190, at 19 (1997) ("Rather than embarking upon a wholesale clarification of these doctrines [of contributory and vicarious liability], the Committee decided to leave current law in its evolving state and, instead, to create a series of 'safe harbors.'"). *Id.* at 45 ("Section 512 does not require use of the notice and take-down procedure. . . . [T]he service provider is free to refuse to 'take down' the material or site, even after receiving a notification . . . in such a situation, the service provider's liability, if any, will be decided without reference to section 512(c).").

192. *See, e.g.*, *Marobie-FL, Inc. v. Nat'l Ass'n of Fire Equip. Distribs.*, 983 F. Supp. 1167 (N.D. Ill. 1997) (dismissing direct and vicarious infringement claims against service provider, and refusing summary judgment on contributory infringement); *Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc.*, 907 F. Supp. 1361, 1377–78 (N.D. Cal. 1995) ("[T]he court is not convinced that Usenet servers are directly liable If Usenet

deemed a question of fact, turning on the degree and timing of notice of alleged infringement, and on their own potential fair use defenses.¹⁹³

Since then, we seem to have reached a scheme of liability on notice. Contributory infringement depends on knowingly materially assisting infringement,¹⁹⁴ and providing hosting to infringing material available for unauthorized download is argued to represent “material assistance.”¹⁹⁵

Absent the DMCA, therefore, service providers would likely not be held liable for infringements of which they were ignorant, nor would they be required to search out infringements after a generalized claim. Depending on circumstances, they could be held liable for contributory infringement if they continued to assist in the transmission of material after being specifically notified of its infringing character, but the specificity of that notice and what would be required to give actual knowledge of infringement might be greater than that required by § 512(c)(3)’s formula, which includes no proof beyond an assertion that the material claimed to have been copied was copyrighted and that its copying was unauthorized.¹⁹⁶

Courts have applied the DMCA safe harbor protections to a variety of Internet sites that host user speech, including Google, Amazon.com, and eBay,¹⁹⁷ as well as to video hosting sites YouTube and Veoh.¹⁹⁸ Google, which receives and responds to notices complaining of links to allegedly infringing material, as a § 512(d) provider of “information location tools,”¹⁹⁹ faces even less likelihood of liability for those hyperlinks, but may consider its visibility as creating a larger potential risk.²⁰⁰ With a goal of indexing all the world’s information,

servers were responsible for screening all messages coming through their systems, this could have a serious chilling effect on what some say may turn out to be the best public forum for free speech yet devised.”). By contrast, service providers had been found liable when they participated in the infringement and earned money directly from it. *See, e.g.,* Sega Enters. Ltd. v. MAPHIA, 948 F. Supp. 923 (N.D. Cal. 1996) (holding liable online bulletin board operators who specifically solicited copying of Sega video games and expressed the desire that Sega video game programs be placed on bulletin board for downloading purposes).

193. *See Marobie-FL*, 983 F. Supp. at 1178–79; *Netcom*, 907 F. Supp. at 1382–83.

194. *See* Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 535 U.S. 913 (2005); A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (9th Cir. 2001).

195. Note that the case for linking liability is even more attenuated. *See* Perfect 10 v. Google, Inc., 416 F. Supp. 2d 828 (C.D. Cal. 2006).

196. 17 U.S.C. § 512(c)(3) (2006).

197. *See* Perfect 10, 416 F. Supp. 2d 828; Corbis Corp. v. Amazon.com, Inc., 351 F. Supp. 2d 1090 (W.D. Wash. 2004); Hendrickson v. eBay, Inc., 165 F. Supp. 2d 1082 (C.D. Cal. 2001).

198. *See* Viacom Int’l Inc. v. YouTube, Inc., Nos. 07 Civ. 2103, 07 Civ. 3582, 2010 U.S. Dist. LEXIS 62829 (S.D.N.Y. June 23, 2010); Io Group, Inc. v. Veoh Network, Inc., 586 F. Supp. 2d 1132 (N.D. Cal. 2008).

199. 17 U.S.C. § 512(d) (2006).

200. Google has been sending the takedown notices it receives to Chilling Effects since 2003, and linking to ChillingEffects.org when results have been removed from a search.

Google will naturally index some information that others would not like to be found. Some complainants will almost certainly be determined and deep-pocketed enough to wage expensive litigation, even if the claims ultimately lack merit.²⁰¹

B. The Chill in Practice

In our non-ideal world, the notice-and-takedown regime has spawned many notices of claimed infringement and many takedowns of allegedly infringing material. In practice, along with expeditious removals of infringing material have come speedy takedowns of non-infringing speech. Additionally, many scenarios simply fall outside the core of copyright's policy justifications, and others are too close to the edge between infringement and fair use to be decided accurately by the summary procedures of a service provider reviewing a § 512(c) notice.

Occasionally, the DMCA has induced flat-out errors. The Recording Industry Association of America ("RIAA") sent a DMCA notice to Penn State's Department of Astronomy and Astrophysics in May 2003, accusing the university of unlawfully distributing songs by the musician Usher, and nearly forcing the department's servers offline during exam period. As it turned out, RIAA had mistakenly identified the combination of the word "Usher" (identifying faculty member Peter Usher) with an a cappella song performed by astronomers about gamma rays as an instance of infringement. In apologizing, RIAA noted that its "temporary employee" had made an error. RIAA admitted that it does not routinely require its "Internet copyright enforcers" to listen to the song that is allegedly infringing.²⁰² In the same period, RIAA admitted to several dozen additional errors in sending accusatory DMCA notices, all made within a single week. But RIAA has refused to provide additional details about these errors, professing concern that to do so would compromise the "privacy" of its employees and of the victims of its false accusations.²⁰³ Sony Mu-

Google puts no conditions on the Chilling Effects publication or analysis of those notices. As of March 2010, Google receives more than a thousand DMCA takedown demands, many citing multiple URLs, each month. The majority of these notices request removal of links from the search index, invoking § 512(d), but a substantial number are § 512(c) notices regarding material Google hosts on its Blogger weblog service or in conjunction with other Google services. *See* CHILLING EFFECTS, <http://www.chillingeffects.org> (last visited Dec. 21, 2010).

201. Perfect 10, a purveyor of naked photographs and lawsuits, has filed thirty copyright lawsuits between January 1999 and September 2010 (PACER search on file with author).

202. *See* Declan McCullagh, *RIAA Apologizes for Threatening Letter*, CNET NEWS (May 12, 2003), http://news.cnet.com/2100-1025_3-1001095.html.

203. *See id.*

sic has been made to retract some notices sent to the recording artists who made the tracks in question and retained copyright therein.²⁰⁴

Likewise, the Internet Archive's historic Prelinger collection of public domain films earned a takedown from Universal Studios over its films' numerical file names. Universal sent a DMCA notice to the Internet Archive in connection with films 19571.mpg and 20571a.mpg, after Universal's bot apparently mistook public domain films on home economics for the copyrighted submarine movie "U-571."²⁰⁵ Because the Internet Archive is a large enough collection to act as its own service provider, it was saved the trouble of explaining this to an upstream service provider who might not have grasped the distinction quickly enough to avoid a shutoff. In a similar case, Warner Brothers threatened a child whose Harry Potter book report wound up in a "shared" folder and was mistaken for the movie.²⁰⁶

The RIAA members' sound recordings and Universal's and Warner's movies are creative works entitled to the full protection of copyright. But the copyright sword against piracy is not supposed to be a blunderbuss. On the face of a DMCA § 512(c) notification,²⁰⁷ there may be little to distinguish innocent from infringing speech, and the legal structure and market pressure give the service provider little incentive to investigate beyond the face of the notice.

Researchers at the University of Washington documented their experience receiving DMCA takedown demands for their networked laser printers, which were not offering for download any of the Iron Man or Indiana Jones movies for which they were accused.²⁰⁸ Instead, the agents sending DMCA notices, and the university service provider passing them along, never verified that files were being offered from the IP address identified in a BitTorrent swarm.

Not all takedowns are commercial or entertainment-related. To those who see the First Amendment primarily as protection for the

204. See Sean Michaels, *Sony Music 'Mistakenly Removed' Bradford Cox Songs*, THE GUARDIAN (London), Nov. 30, 2010, <http://www.guardian.co.uk/music/2010/nov/30/sony-music-bradford-cox-songs>.

205. See *Universal Studios Stumbles on Internet Archive's Public Domain Films*, CHILLING EFFECTS (Feb. 27, 2003), <http://www.chillingeffects.org/notice.cgi?NoticeID=595>.

206. See Roy Mark, *Verizon Seeks Stay of RIAA Ruling*, INTERNET NEWS (Jan. 30, 2003), <http://www.internetnews.com/bus-news/article.php/1577111>.

207. 17 U.S.C. § 512(c) (2006).

208. See Michael Piatek, Tadayoshi Kohno & Arvind Krishnamurthy, *Challenges and Directions for Monitoring P2P File Sharing Networks — or — Why My Printer Received a DMCA Takedown Notice*, HOTSEC '08, http://www.usenix.org/event/hotsec08/tech/full_papers/piatek/piatek.html (last visited Dec. 21, 2010); Brad Stone, *The Inexact Science Behind D.M.C.A. Takedown Notices*, N.Y. TIMES BITS (June 5, 2008, 11:18 AM), <http://bits.nytimes.com/2008/06/05/the-inexact-science-behind-dmca-takedown-notices/>.

political discourse necessary for self governance,²⁰⁹ a range of politically oriented takedowns should raise concern. Some, such as the presidential campaign videos described in the introduction,²¹⁰ or the Diebold email archives described below,²¹¹ play out on a grand stage; others are more local.

The New York State College Republicans, amid a contested battle for control of the College Republicans organization, sent a takedown notice against the weblog “Musings of a New York College Republican,” alleging that it copied several photographs and “engaged in ‘remote loading’” of several press releases.²¹² The anonymous blogger had been critical of infighting in the College Republicans organization.²¹³ The senders of the demand requested identification of the anonymous blogger and threatened legal action if they did not receive it.²¹⁴ The New York State College Republicans sent the DMCA takedown notice even though “remote loading” is simply hyperlinking to a page on a different server — something that web pages do every day — and highly unlikely to be found to constitute copyright infringement. Indeed, “remote loading” is an alternative to copying the content to which you want to make reference.

A graphic designer sent a DMCA complaint when the conservative Arkansas Family Coalition weblog used an Arkansas Democrat’s campaign logo to illustrate a post discussing ethics complaints regarding campaign contributions accepted by the candidate.²¹⁵ Although a political logo can be copyrighted like any other graphic design, there

209. See generally ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT (1948).

210. See *supra* Part I.

211. See *infra* Part IV.D. The Center for Democracy and Technology documents numerous copyright takedowns affecting political candidates in a recent report. See CTR. FOR DEMOCRACY & TECH., CAMPAIGN TAKEDOWN TROUBLES: HOW MERITLESS COPYRIGHT CLAIMS THREATEN ONLINE POLITICAL SPEECH (2010), http://www.cdt.org/files/pdfs/copyright_takedowns.pdf. Other large-scale examples include the National Organization for Marriage, whose video got a rare fair use review after DMCA takedown and was reposted before ten business days had elapsed. See Sam Bayard, *YouTube Restores National Organization for Marriage Video Outside DMCA Parameters, Cites Fair Use*, CITIZEN MEDIA LAW PROJECT (May 7, 2009), <http://www.citmedialaw.org/blog/2009/youtube-restores-national-organization-marriage-video-outside-dmca-parameters-cites-fair-u>.

212. See *NY College Republicans Complain About Critics*, CHILLING EFFECTS (July 22, 2005), <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=2174>.

213. The original posts are unavailable, but some of the debate is visible in slightly later-archived posts. See, e.g., *Let’s Do This*, MUSINGS OF A NEW YORK COLLEGE REPUBLICAN (Oct. 7, 2005, 2:08 AM), <http://web.archive.org/web/20051215011122/http://nycr.blogspot.com/2005/10/lets-do-this.html> (preserved at the Internet Archive).

214. See *id.*

215. See *Graphic Designer Complains of Use of Political Logo*, CHILLING EFFECTS (Oct. 7, 2005), <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=2455>. The logo was originally included as a graphic at *Jimmy Lou Fisher Facing Ethics Complaint over Illegal Campaign Contributions*, Arkansas Family Coalition- ArkFam.com (Oct. 7, 2005), <http://web.archive.org/web/20061029112839/arkansasfamilycoalition.blogspot.com/2005/10/jimmie-lou-fisher-facing-ethics.html> (preserved at the Internet Archive).

is a strong argument that fair use permits commentary that uses a candidate's political logo to identify the candidate.

Photographer Leif Skoogfors sent numerous DMCA complaints when a photograph he had taken at a 1970 Vietnam peace rally, showing Jane Fonda in the foreground and John Kerry behind, appeared on anti-Kerry sites around the web.²¹⁶ It was debatable whether the image contradicted what Senator Kerry, then a presidential candidate, was claiming about his Vietnam-era opposition to the war or whether the image was being misrepresented and blown out of proportion. But even the photographer Skoogfors acknowledged, "Now the picture was the news."²¹⁷ People on both sides of the political discussion "quoted" the photograph to support their arguments. Yet when service providers received DMCA takedown notices, most removed the pictures.

Activists The Yes Men saw how far copyright could reach when they criticized the Dow Chemical Company by taking a copy of Dow's website and creating one at dow-chemical.com that apologized for chemical accidents at Bhopal. As The Yes Men designed, Dow then had to disavow the apology, a move The Yes Men took as a renewed opportunity for criticism.²¹⁸

Dow further responded with a DMCA takedown complaint to Verio, the owner of the netblock in which the dow-chemical.com site was hosted, which stated that "[t]he Website displays numerous trademarks, images, texts and designs taken directly from Dow's website located at dow.com. This material is protected by copyright law and may not be reproduced, in whole or in part, without the express written authorization of Dow."²¹⁹ The Yes Men's hosting provider, New York service provider Thing.net, indicated it would not take down the material, but the target of Dow's letter was one level up the chain. Verio, provider of connectivity and network space to Thing.net, was not swayed by Thing.net's determination to stand by its custom-

216. See Search for: "skoogfors," CHILLING EFFECTS, <http://www.chillingeffects.org/search.cgi?search=skoogfors> (last visited Dec. 21, 2010). It appears that the Skoogfors photograph was generally presented in its original form, with occasional labels marking the figures. Another photograph, circulated at the same time, was doctored to place Kerry and Fonda on the same podium. See Ken Light, Editorial, *Fonda, Kerry and Photo Fakery*, WASH. POST, Feb. 28, 2004, at A21. At the same time, it does not appear that Skoogfors was trying to get his photograph withdrawn from debate entirely — it continued to be available for licensing from Corbis. See Leif Skoogfors, *A Face in the Crowd*, THE DIGITAL JOURNALIST (Mar. 2004), http://www.digitaljournalist.org/issue0403/dis_skoogfors.html.

217. Skoogfors, *supra* note 216.

218. See *Dow Hijink*, THE YES MEN, <http://web.archive.org/web/20050527011009/http://www.theyesmen.org/hijinks/dow/bhopal2002.shtml> (last visited Dec. 21, 2010) (preserved at the Internet Archive).

219. See Letter from Gregory D. Phillips, Howard, Phillips & Andersen, to Verio, Inc. (Dec. 3, 2002), available at http://web.archive.org/web/20050511184502/www.theyesmen.org/hijinks/dow/Dow-Chemical_DMCAnotice.pdf (preserved at the Internet Archive).

ers. In the ensuing scuffle, Verio cut connectivity to all of Thing.net's customers for twelve hours.²²⁰

The Yes Men's parody²²¹ might or might not have crossed the line from parody to copyright infringement — as well as trademark infringement and false advertising — but nothing was alleged against the other digital artists who hosted sites with Thing.net. The Yes Men have engaged in several further online parodies and protests, prompting repeat uses of the DMCA.²²²

Further, if we do not presume that everyone in a political debate will act civilly — and one of the reasons for constitutional government and its procedures is precisely to restrain us when we act uncivilly — we should be wary of mechanisms that give one party the ability to shut down debate rather than participate in it. Especially in political debate, one often wants to quote from one's opponent. If every such quotation brings threat of a facially plausible copyright takedown, the scope of political debate is narrowed. When some of those threatened claims materialize, it is narrowed further.

The DMCA proves attractive to those looking to take down speech they find annoying — precisely the kind of speech that may be of most interest to political debate. The website cryptome.org (“Cryptome”) has a history of publishing leaked documents, from the DeCSS DVD decryption code²²³ to unredacted and incorrectly redacted versions of TSA screening documents.²²⁴ Various companies and government agencies have requested that the site or its pages be taken down, but few of them have succeeded.²²⁵ Recently, Cryptome has been collecting the surveillance price lists from various Internet com-

220. See *Routledge Just Says “Yes” to Dow: The Collaboration of a Progressive Academic Press and a Large Chemical Corporation*, THE YES MEN, <http://theyesmen.org/dowtext/> (last visited Dec. 21, 2010).

221. After losing control of the dow-chemical.com domain, The Yes Men moved their spoof Dow site to <http://dowethics.com/>. See *id.*

222. See, e.g., Wendy Davis, *ISP Takes Down Parody After Chamber of Commerce Complaints*, MEDIAPOST (Oct. 25, 2009), http://www.mediapost.com/publications/?fa=Articles.showArticle&art_aid=116054; *NYT spoof*, DIAGONAL THOUGHTS (Nov. 12, 2008, 10:00 PM), <http://www.diagonalthoughts.com/?p=397> (reporting on a New York Times parody website that was taken down after DeBeers complained of a fake advertisement on the site). The Yes Men maintain a list of their latest “hijinks” at <http://theyesmen.org/hijinks>.

223. See *MPAA Notice to Cryptome on DeCSS*, CRYPTOME, <http://cryptome.org/dvd-mpaa-cdd.htm> (last visited Dec. 21, 2010).

224. See *TSA Blows Smoke for Sensitive Screening Document*, CRYPTOME, <http://cryptome.org/tsa-smoke/tsa-smoke.htm> (last visited Dec. 21, 2010).

225. Yahoo sent a DMCA takedown notice directly to Cryptome after they published Yahoo's information price list. See *Yahoo Tries To Hide Snoop Service Price List*, ELEC. FRONTIER FOUND., <http://www.eff.org/takedowns/yahoo-tries-hide-snoop-service-price-list> (last visited Dec. 21, 2010); Kim Zetter, *Yahoo Issues Takedown Notice for Spying Price List*, WIRED THREAT LEVEL (Dec. 4, 2009, 5:00 PM), <http://www.wired.com/threatlevel/2009/12/yahoo-spy-prices/>.

panies.²²⁶ Microsoft expressed its dissatisfaction to this practice in a DMCA notice to Cryptome's web host and domain name registrar, Network Solutions:

Microsoft has received information that the domain listed above, which appears to be on servers under your control, is offering unlicensed copies of, or is engaged in other unauthorized activities relating to copyrighted works published by Microsoft.

1. Identification of copyrighted works:

Copyrighted work(s):

Microsoft Global Criminal Compliance Handbook²²⁷

Accordingly, Network Solutions notified Cryptome's proprietor, John Young, that it would have to disable the entire site for the ten-business-day period unless he removed the page. Young counter-notified but refused to remove the page, and so, despite his assertions of fair use, Network Solutions deactivated the entire site — the only way they believed they could comply with the DMCA.²²⁸ After the buzz of publicity, so common an occurrence it has been named the "Streisand Effect,"²²⁹ kicked in, Microsoft retracted its DMCA complaint, enabling Network Solutions to restore Cryptome.²³⁰

Other claims misinterpret the scope of copyright exclusivity. The Church of Scientology was a pioneer in using the DMCA to ask Google to de-index websites critical of the Church, on the grounds that the criticism on the websites quoted from Scientology texts.²³¹

226. See *Online Spying Guides*, CRYPTOME, <http://cryptome.org/isp-spy/online-spying.htm>.

227. See *Microsoft Demands Takedown of Microsoft Spy Guide*, CRYPTOME, <http://cryptome.org/0001/ms-spy-takedown.htm> (last visited Dec. 21, 2010).

228. See *id.*

229. The Streisand Effect is named for a 2003 incident in which Barbra Streisand sued a California photographer for including aerial photographs of her Malibu house on his coastal survey website. Instead of removing the image, the photographer publicized the suit, drawing further attention to the photo. See Andy Greenberg, *The Streisand Effect*, FORBES, May 11, 2007, http://www.forbes.com/2007/05/10/streisand-digg-web-tech-cx_ag_0511streisand.html. Adelman, who maintained californiacoastline.org, obtained dismissal of the suit under California's anti-SLAPP law, and won attorneys' fees and costs. See *Streisand v. Adelman*, No. SC 077 257 (L.A. Sup. Ct. May 10, 2004) (order granting in part and denying in part plaintiff's motion to tax costs and defendants' motions for attorneys' fees), available at <http://www.californiacoastline.org/streisand/fees-ruling.pdf>.

230. See Chloe Albanesius, *Cryptome Restored After Microsoft DMCA Takedown*, PCMAG.COM, Feb. 25, 2010, <http://www.pcmag.com/article2/0,2817,2360694,00.asp>.

231. See *Google Asked To Delist Scientology Critics (#1)*, CHILLING EFFECTS (Mar. 8, 2002), <http://www.chillingeffects.org/notice.cgi?NoticeID=232>; *Takedown Demands*, CHILLING EFFECTS (Mar. 8, 2002); see also Matt Loney & Evan Hansen, *Google Pulls Anti-Scientology Links*, CNET NEWS (Mar. 21, 2002), <http://news.com.com/2100-1023-865936.html> (detailing use of DMCA in conflict between Church of Scientology and anti-Scientology website).

Quotation for the purpose of criticism and commentary is a specifically designated fair use,²³² highly likely to be found “transformative.”²³³

Wal-Mart sent a § 512(h) subpoena, along with a § 512(c) notice, to a comparison-shopping website that allowed customers to post prices of items sold in stores, claiming incorrectly that its prices were copyrighted. Wal-Mart sought the identity of the user who had anonymously posted information about an upcoming sale. Other retailers, including Kmart, Jo-Ann Stores, OfficeMax, Best Buy, and Staples, also served § 512(c) notices on the website based on the same theory of copyrightable facts.²³⁴ While they might have had trade secret misappropriation claims against those who leaked circulars before holiday sales (and, less plausibly, a claim that the websites should have known the information was misappropriated), asking a judge for a temporary restraining order would have required more time, money, and effort than simply sending DMCA notices to the service providers.

Finally, some claims use the DMCA as a battering ram, seemingly assuming that where text exists, so too does copyright infringement if one looks hard enough.

Mir Internet Marketing offers “full-service, cost-effective, end-to-end Internet marketing solutions,” including search engine optimization.²³⁵ Their service, in short, is to get clients’ websites to appear in response to searches on favored keywords, aiming to maximize the number of searchers who click through to the clients’ sites.²³⁶ Mir and other optimizers have added another trick to their bags — DMCA takedowns against competitors. After all, removing competing pages from search engine results boosts the visibility of your remaining sites.

232. See 17 U.S.C. § 107 (2006) (“The fair use of a copyrighted work, including such use by reproduction in copies . . . for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright.”).

233. See *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994) (holding a music group’s parody to be a form of criticism, likely fair use).

234. See Declan McCullagh, *Wal-mart Backs Away from DMCA Claim*, CNET NEWS (Dec. 5, 2002), <http://news.cnet.com/2100-1023-976296>; *Press Release: FatWallet Challenges Abusive DMCA Claims*, FATWALLET (Dec. 2, 2002, 4:25 PM), <http://www.fatwallet.com/forums/messageview.cfm?catid=18&threadid=129657>.

235. Mir Internet Marketing Homepage, <http://www.internetmadeeasy.com/> (last visited Dec. 21, 2010).

236. The tactics of search engine optimizers vary. One tactic is to optimize the site structure for search engine crawlers by including common search terms in text and links. Another is to link to the site from other high-traffic pages (whether pages with real relevance or fake sites designed solely to generate “link rank”). An additional tactic is to “farm” links out through spam or on typo-sites. Finally, some search engine optimizers bury keywords in hidden text or “gateway pages.” See *Search Engine Optimization*, WIKIPEDIA, http://en.wikipedia.org/wiki/Search_engine_optimization (last visited Dec. 21, 2010).

Mir has sent at least forty-eight separate takedown notices against hundreds of websites it claims infringe its copyrights, often based on a few duplicate phrases.²³⁷ In a section of its website describing the DMCA takedown process, Mir (speaking through its SEO Logic division) says, “We consider removing violators to be part of our job in helping our clients to improve their search engine ranking.”²³⁸

DMCA takedowns may target non-infringing as well as infringing uses. When competitors choose to “borrow” substantial text someone else has written rather than writing their own, they infringe copyright. Often, however, similarities reflect not direct copying but the relatively limited number of ways to describe a generic product or service.²³⁹ For example, much of the text on many advertising sites is minimally creative recitation of fact. Also, dozens of the DMCA takedown notices between competitors in the Chilling Effects archives target insubstantial similarities.²⁴⁰ Because there are only so many ways to describe a generic product or service, the uses of the copyrighted material indicated in the results of a search for that material are not necessarily infringing.

The chief concern of search engine optimizers is not to remedy the infringement, but to penalize competitors. Mir therefore recommends that copyright holders contact search engines and the site’s service provider before contacting the allegedly infringing site’s webmaster:

Do not contact the owner or Webmaster of the site
that is illegally using your content. . . . If you want to

237. See Search for: “Mir Internet Marketing,” CHILLING EFFECTS, <http://www.chillingeffects.org/search.cgi?search=%22Mir+Internet+Marketing%22> (last visited Dec. 21, 2010).

238. Search Engine Marketing FAQ: Copyright Infringement and DMCA, SEO LOGIC, <http://www.seologic.com/faq/copyright.php> (last visited Dec. 21, 2010). Mir even advises clients that they can “see examples of our DMCA filings with Google, [by] visit[ing] ChillingEffects.org, where Google posts copies of all notices it receives.” Search Engine Marketing FAQ: Sending DMCA Notifications, SEO LOGIC, <http://www.seologic.com/faq/dmca-notifications.php> (last visited Dec. 21, 2010).

239. The merger doctrine permits copying when there are so few ways of expressing an idea that protecting the expression would grant monopoly on the idea. See, e.g., N.Y. Mercantile Exch., Inc. v. IntercontinentalExchange, Inc., 497 F.3d 109, 117 (2d Cir. 2007) (holding exchange’s settlement prices merged with their ideas). The related doctrine of scenes a faire excuses the use of stock scenes or phrases “if the expression embodied in the work necessarily flows from a commonplace idea.” Ets-Hokin v. Skyy Spirits, Inc., 225 F.3d 1068, 1082 (9th Cir. 2000) (discussing backlit photograph of vodka bottle against a plain background).

240. Takedown notices sent by competitors are compiled at *Competition*, CHILLING EFFECTS, <http://www.chillingeffects.org/dmca512/keyword.cgi?KeywordID=36> (last visited Dec. 21, 2010). Because the letters themselves do not include the full text of the original website or the identity of the alleged infringer, and the sites’ content may have changed since the letters were sent, after-the-fact comparison will not be foolproof. The same factors make it difficult for the hosting service provider or search engine to evaluate DMCA infringement claims.

punish the Webmaster for copying your content, and have their site removed from the search engines, or even from the Internet entirely, then you should take the following steps . . .

File notices of alleged infringement that comply with the Digital Millennium Copyright Act (DMCA) with each search engine or directory where the infringing site is listed. The Digital Millennium Copyright Act empowers you to send a notice to any directory or search engine that lists the offending site and demand that they remove any links to the offending site. Yes, you can make Google, Yahoo!, and all the others take the site out of their search results.²⁴¹

This practice appears to have sprung up in direct response to the DMCA, specifically driven by § 512(d)'s instructions to providers of "information location tools."²⁴²

In the rush for page-views, some of those looking for advantage will skirt the law. What they want is precisely what the DMCA induces search engines to offer — rapid unquestioning takedown, for at least a short period of time. Copyright takedowns serve as tools in competitive scrambles for attention,²⁴³ partnership disputes,²⁴⁴ and disputes between independent contractors and their clients.²⁴⁵

C. "Repeat Infringers"

When a number of music blogs disappeared from Google's Blogger service, where they were hosted, their authors found entire sites, sometimes including years of archives, deleted. The bloggers were notified that "Upon review of your account, we've noted that your blog has repeatedly violated Blogger's Terms of Service . . . [and]

241. *Search Engine Marketing FAQ: Copyright Infringement and DMCA*, SEO LOGIC, <http://www.seologic.com/faq/copyright.php> (last visited Dec. 21, 2010).

242. 17 U.S.C. § 512(d) (2006).

243. See, for example, the dispute between two makers of chef's jackets, Crooked Brook and Bragard, the former complaining about a quotation posted on the latter's website. *Crooked Brook Complains About Copied Chef Coat Text*, CHILLING EFFECTS (May 2, 2006), <http://www.chillingeffects.org/notice.cgi?sID=1359>.

244. Hosting companies have been asked to remove or disable access to web pages pursuant to the DMCA in disputes between former partners over ownership of jointly created content. See, e.g., *Golden Gate Expeditions Complaint to Web Host*, CHILLING EFFECTS (Feb. 20, 2003), <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=572>.

245. Google was asked to remove links to web pages pursuant to the DMCA in disputes between an independent contractor and its client over the ownership of a website the contractor designed but for which he alleged he was not paid. *Azalea Web Design Company Asks Google to Delist Client*, CHILLING EFFECTS (May 2, 2004), <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=1256>.

we've been forced to remove your blog."²⁴⁶ According to reports, the bloggers had run afoul of Google's "repeat infringer" policy²⁴⁷ after their blogs were the subject of several complaints.²⁴⁸ Pursuant to that policy, inspired by the DMCA's requirement,²⁴⁹ Google opted to terminate their accounts, removing not only the allegedly infringing entries, but the entirety of the blogs' content.²⁵⁰

While some "blogs," on Blogger and elsewhere, appeared to be mere collections of links to newly released songs and albums, others, including some taken down in the "music blogocide,"²⁵¹ were written by music critics who linked to songs in order to enhance their commentary or to alert readers to new music.²⁵² Some blog authors asserted that they operated with the permission, or even the encouragement, of the artists or music labels whose work they posted.²⁵³

Copyright claimants urge that two or three takedown notices make someone a "repeat infringer" whose account must be terminated. In contrast, David Nimmer suggests that the provision should be construed strictly, to require "repeat infringer" sanctions only against those who have more than once been found liable for copyright infringement after legal proceedings.²⁵⁴ Taking a middle course,

246. Sean Michaels, *Google Shuts Down Music Blogs Without Warning*, THE GUARDIAN (London), Feb. 11, 2010, <http://www.guardian.co.uk/music/2010/feb/11/google-deletes-music-blogs>.

247. See *Digital Millennium Copyright Act — Blogger*, GOOGLE, http://www.google.com/blogger_dmca.html (last visited Dec. 21, 2010) ("Many Google Services do not have account holders or subscribers. For Services that do, such as Blogger, Google will, in appropriate circumstances, terminate repeat infringers.").

248. See CHILLING EFFECTS, <http://www.chillingeffects.org/> (a search for "irockcleveland" returns seven notices from IFPI between August 2009 and February 2010).

249. 17 U.S.C. § 512(i)(1)(A) (2006) states:

The limitations on liability established by this section shall apply to a service provider only if the service provider (A) has adopted and reasonably implemented, and informs subscribers and account holders of the service provider's system or network of, a policy that provides for the termination in appropriate circumstances of subscribers and account holders of the service provider's system or network who are repeat infringers.

250. See Michaels, *supra* note 246.

251. The tag "#Musicblogocide2k10" reached the trend charts on Twitter when Google removed many music blogs from Blogger. See Michaels, *supra* note 246.

252. See, for example, the relocated I ROCK CLEVELAND, <http://blog.irockcleveland.com> (last visited Dec. 21, 2010).

253. See Scott Jagow, *Google commits "Blogocide,"* SCRATCH PAD (Feb. 11, 2010, 1:15 PM), http://www.publicradio.org/columns/marketplace/scratchpad/2010/02/google_commits_blogocide.html (quoting the owner of music blog I Rock Cleveland as writing back to Google, "I assure you that everything I've posted for, let's say, the past two years, has either been provided by a promotional company, came directly from the record label, or came directly from the artist").

254. David Nimmer, *Repeat Infringers*, 52 J. COPYRIGHT SOC'Y U.S.A. 167, 195–98 (2005). Nimmer also notes that unless the imposition of strikes is discretionary rather than mandatory, all of the major motion picture studios would be ineligible for online posting

Google has chosen to remove a blog “when [it] receives multiple DMCA complaints about the same blog, and [has] no indication that the offending content is being used in an authorized manner.”²⁵⁵

The due process afforded by the takedown and termination process is insufficient, particularly given the severity of the process. Many bloggers never realized that merely removing entries on which they had received complaints was not sufficient to clear their records. At least one of the February removals was reinstated after Google admitted that notifications of the prior DMCA complaints had failed to reach the blogger.²⁵⁶ Many of the IFPI notices to Google in the Chilling Effects database²⁵⁷ lack basic elements of the DMCA § 512(c)(3)(A) notification,²⁵⁸ including “[i]dentification of the copyrighted work claimed to have been infringed,”²⁵⁹ and “[i]dentification of the material that is claimed to be infringing,”²⁶⁰ as they list only URLs to posts, not to the linked files.²⁶¹ It appears that IFPI claims that, as a U.K.-based organization, it need not meet the U.S. DMCA requirements, and that Google has chosen not to press the point.²⁶²

This Article does not claim that all of the above examples represent clear-cut cases of non-infringement. The uses are not necessarily fair and non-infringing; the senders of takedown notices are not nec-

accounts, since all have had multiple copyright infringement judgments rendered against them. *Id.* at 216–17.

255. Rick Klau, *A Quick Note About Music Blog Removals*, *BLOGGER BUZZ* (Feb. 10, 2010, 2:31 PM), <http://buzz.blogger.com/2010/02/quick-note-about-music-blog-removals.html>; see also Wendy Seltzer, *DMCA “Repeat Infringers”: Scientology Critic’s Account Reinstated after Counter-Notification*, *CHILLING EFFECTS* (June 6, 2008), <http://www.chillingeffects.org/weather.cgi?WeatherID=605> (chronicling Scientology critic’s experience with the DMCA and Google subsidiary YouTube).

256. See Klau, *supra* note 255; *Musicblogocide2k10: La Vie Continue*, *MASALACISM* (Feb. 12, 2010), <http://www.masalacism.com/2010/02/musicblogocide2k10-la-vie-continue/>.

257. See *Search for “IFPI,” CHILLING EFFECTS*, <http://www.chillingeffects.org/search.cgi?q=IFPI> (last visited Dec. 21, 2010) (listing notices).

258. See 17 U.S.C. § 512(c)(3)(A) (2006).

259. *Id.* § 512(c)(3)(A)(ii).

260. *Id.* § 512(c)(3)(A)(iii).

261. See, e.g., *IFPI DMCA (Copyright) Complaint to Google*, *CHILLING EFFECTS* (Mar. 12, 2010), <http://www.chillingeffects.org/dmca512c/notice.cgi?NoticeID=33815> (“We have learned that your service is hosting the above web sites on your network. *These sites are offering direct links to files* containing sound recordings for other users to download. The copyright in these sound recordings is owned or exclusively controlled by certain IFPI Represented Companies.”) (emphasis added). According to the DMCA:

[A] notification from a copyright owner or from a person authorized to act on behalf of the copyright owner that fails to comply substantially with the provisions of subparagraph (A) shall not be considered . . . in determining whether a service provider has actual knowledge or is aware of facts or circumstances from which infringing activity is apparent.

17 U.S.C. § 512(c)(3)(B)(i) (2006).

262. See Stelios Philis, *Reinvestigating Music Blogocide 2k10: Google is Less Evil than We Think*, *POPSense* (Feb. 22, 2010), <http://www.popsense.com/2010/02/reinvestigating-music-blogocide-2k10.html>.

essarily motivated by invidious purposes. The claim is rather that neither are they clear-cut cases of infringement. In equivocal cases, of which copyright has many, the benefit of the doubt should lie with the speaker. Instead, the summary process of takedown upon DMCA notice to a third party deprives parties, the public, and the law of an important opportunity to clarify.²⁶³

Nor does this Article claim that a majority of takedowns are improper.²⁶⁴ It is likely that many people posting copyrighted music or movie files in their entirety have no non-infringing purpose and no objective other than avoiding payment for a commercially available work. Some posters of others' images and text will have no fair use or other defenses. At the same time, this Article has identified only a few of the many erroneous takedowns.²⁶⁵ The argument here does not depend on proportions; the volume of infringement does not excuse a regime systematically vulnerable to speech-chilling errors.

Elsewhere, Rebecca Tushnet analogizes copyright restriction to a poll tax or literacy test setting discriminatory barriers to expression.²⁶⁶ The DMCA notification and counter-notification regime, even if ultimately navigable, poses similar hurdles. As the Court said of book-sellers in *Smith v. California*, "The [service provider's] self-censorship, compelled by the State, would be a censorship affecting the whole public, hardly less virulent for being privately administered."²⁶⁷

D. Limited Warming?

The DMCA includes a provision, § 512(f), that allows the targets of improper takedowns to file suit against the takedown senders.²⁶⁸ A

263. As this Article goes to press, another round of music blog takedowns has occurred, this time through the seizure of domain names by the department of Immigration and Customs Enforcement, working with the RIAA. Again, some bloggers assert that they acted with permission from copyright holders. See Ben Sisario, *Piracy Fight Shuts Down Music Blogs*, N.Y. TIMES, Dec. 14, 2010, at B1, available at <http://www.nytimes.com/2010/12/14/business/media/14music.html>.

264. I do note, following Urban and Quilter, that many ostensible DMCA notices demanding takedown do not comply with even the minimal requirements of § 512(c)(3). See Urban & Quilter, *supra* note 160, at 667–68 (finding a substantial percentage of notices suffered from substantive or formal defects).

265. Others are described at Chilling Effects and the EFF's "No Downtime for Free Speech" Campaign and Takedown Hall of Shame. See *No Downtime for Free Speech Campaign*, ELEC. FRONTIER FOUND., <http://www.eff.org/issues/ip-and-free-speech> (last visited Dec. 21, 2010); *Takedown Hall of Shame*, ELEC. FRONTIER FOUND., <http://www.eff.org/takedowns> (last visited Dec. 21, 2010).

266. See Library of Congress Rulemaking Hearing on Section 1201 (2009) (comments of Professor Rebecca Tushnet), available at <http://www.copyright.gov/1201/hearings/2009/transcripts/1201-5-7-09.txt>.

267. 361 U.S. 147, 153–54 (1959).

268. 17 U.S.C. § 512(f) (2006).

few cases under that provision have produced a limited warming effect.

In July 2003, an archive of email messages was leaked from Diebold, Inc., a manufacturer of electronic voting machines. The archive included communications among Diebold employees and contractors describing flaws, sham test messages, and use of uncertified code in electronic voting machines deployed around the country. In October, wanting to share the evidence with others — and to get help reviewing the thousands of messages in the archives for more examples — journalists and activists posted the archive online and invited others to search and mirror the collection. As quickly as mirror sites and search tools were built, Diebold responded with dozens of takedown notices alleging that the postings, and even sites linking to the postings, violated Diebold copyrights. Service providers, including colleges and universities, pulled the web pages. Thus, shortly before the November 2003 elections, many service providers silenced sites discussing voting security.²⁶⁹

The creativity in these e-mails was more in the fudged demonstrations and certifications they described than in the expression copyright protects.²⁷⁰ If anything, the technical details of machine function and malfunction might be a subject for trade secret, rather than copyright. But because “the DMCA provides the rapid response, the rapid remedies that Congress had in mind,”²⁷¹ and a route through service providers and not individuals, Diebold chose to assert copyright infringement rather than trade secret misappropriation.

Two Swarthmore College students had their website disrupted just as they were planning a symposium on the security of electronic voting. Their college, which was hosting the student group’s site,

269. See Declaration of Wendy Seltzer in Support of Plaintiffs’ Application for Temporary Restraining Order at 4, *Online Policy Group v. Diebold, Inc.*, 337 F. Supp. 2d 1195 (N.D. Cal. 2004) (No. C 03-4913 JF), available at http://www.eff.org/files/filenode/OPG_v_Diebold/Seltzer.pdf; Kim Zetter, *E-Vote Protest Gains Momentum*, WIRED.COM (Oct. 29, 2003), <http://www.wired.com/politics/law/news/2003/10/61002>.

270. See *Targeting Diebold with Electronic Civil Disobedience*, WHY WAR?, <http://why-war.com/features/2003/10/diebold.html> (last visited Dec. 21, 2010). The emails stated:

For a demonstration I suggest you fake it. Program [sic] them both so they look the same, and then just do the upload from [sic] the AV. That is what we did in the last AT/AV demo.

. . . .

I have become increasingly concerned about the apparent lack of concern over the practice of writing contracts to provide products and services which do not exist and then attempting to build these items on an unreasonable timetable with no written plan, little to no time for testing, and minimal resources. It also seems to be an accepted practice to exaggerate our progress and functionality to our customers and ourselves then make excuses at delivery time when these products and services do not meet expectations.

Id.

271. *Online Policy Group*, 337 F. Supp. 2d at 1204 n.15.

chose to follow the DMCA's takedown procedure when it received notice from Diebold, notwithstanding letters from the students' counsel.²⁷² Online Policy Group ("OPG"), a non-profit service provider, resisted the takedown demand aimed at a co-located IndyMedia website that linked to the Diebold archive, only to find its upstream service provider threatened with litigation for hosting the intransigent OPG.²⁷³

At this point, OPG, the Swarthmore students, and their pro bono counsel filed suit for DMCA misuse, claiming that Diebold's takedown notices "knowingly materially misrepresented" copyright infringement in violation of § 512(f).²⁷⁴ After suit was filed, Diebold attempted to moot the lawsuit by withdrawing its threats,²⁷⁵ perhaps because Diebold recognized its legal error and that litigation would only serve to bring more attention to the archives and their contents.

The district court granted summary judgment to the plaintiffs, finding that "[t]he email archive was posted or hyperlinked to for the purpose of informing the public about the problems associated with Diebold's electronic voting machines," which made at least a portion of the posting fair use, not infringement as alleged.²⁷⁶

No reasonable copyright holder could have believed that the portions of the email archive discussing possible technical problems with Diebold's voting machines were protected by copyright, and there is no genuine issue of fact that Diebold knew — and indeed that it specifically intended — that its letters to OPG and Swarthmore would result in prevention of publication of that content. The misrepresentations were material in that they resulted in removal of the content from websites and the initiation of the present lawsuit. The fact that Diebold never actually brought suit against any alleged infringer suggests strongly that Diebold sought to use the DMCA's safe harbor provisions — which were designed to protect

272. See *id.* at 1198; Complaint at 13, *Online Policy Group*, 337 F. Supp. 2d 1195 (No. C 03-04913 JF), available at http://www.eff.org/legal/ISP_liability/OPG_v_Diebold/complaint.php; Declaration of Vincent V. Carissimi Regarding Plaintiffs' Application for Preliminary Injunction at 2, *Online Policy Group*, 337 F. Supp. 2d 1195 (No. C 03-04913 JF), available at http://www.eff.org/files/filenode/OPG_v_Diebold/reply_decl_carissimi.pdf.

273. See *Online Policy Group*, 337 F. Supp. 2d at 1198.

274. See Complaint at 13, *Online Policy Group*, 337 F. Supp. 2d 1195 (No. C 03-04913 JF), available at http://www.eff.org/legal/ISP_liability/OPG_v_Diebold/complaint.php. The author was a member of the Electronic Frontier Foundation legal team representing OPG. Stanford's Center for Internet & Society represented the Swarthmore students.

275. See *Online Policy Group*, 337 F. Supp. 2d at 1202.

276. *Id.* at 1203.

[service providers], not copyright holders — as a sword to suppress publication of embarrassing content rather than as a shield to protect its intellectual property.²⁷⁷

In the wake of this ruling, Diebold settled with the plaintiffs for \$125,000.²⁷⁸

In the meantime, however, the DMCA had turned a copyright claim too weak to withstand summary judgment into an instrument of widespread takedown. Diebold's claims and the service providers' prompt resort to the safe harbor resulted in the removal of this non-infringing contribution to political debate from most places on the Internet. Even those who filed counter-notifications had their speech suppressed during critical pre-election days. For those without counsel, this first step, takedown, would likely also be the last.

In *Lenz v. Universal*,²⁷⁹ the same court followed its *Online Privacy Group v. Diebold* ruling with a series of rulings bolstering § 512(f). Stephanie Lenz had posted a short video of her infant son dancing to Prince's "Let's Go Crazy," which could be heard faintly in the background. Universal sent YouTube a takedown that resulted in the video's removal. In addition to counter notifying, Lenz filed suit. The court held that consideration of possible fair use claims was a necessary part in the sending of a valid takedown: "The DMCA . . . requires copyright holders to make an initial review of the potentially infringing material prior to sending a takedown notice A consideration of the applicability of the fair use doctrine is simply part of that initial review."²⁸⁰

The *Lenz* court recognized the public speech interest involved in DMCA takedowns: "the unnecessary removal of non-infringing material causes significant injury to the public where time-sensitive or controversial subjects are involved and the counter-notification remedy does not sufficiently address these harms."²⁸¹ More recently, the court gave victims of abusive takedowns a legal interpretation that would help to vindicate that public interest, holding that 512(f) entitled the target of a misfired takedown to file suit even if the plaintiff's only damages were non-pecuniary.²⁸²

Still, the reach of 512(f) is limited. *Rossi v. Motion Picture Assoc. of Am.* cabins the applicability of this cause of action by emphasizing

277. *Id.* at 1204–05.

278. See *Online Policy Group v. Diebold*, ELEC. FRONTIER FOUND., <https://www EFF.ORG/cases/online-policy-group-v-diebold> (last visited Dec. 21, 2010).

279. *Lenz v. Universal Music Corp.*, 572 F. Supp. 2d 1150 (N.D. Cal. 2008).

280. *Id.* at 1155.

281. *Id.* at 1156.

282. See *Lenz v. Universal Music Corp.*, No. C 07-3783 JF, 2010 WL 702466 at *12 (N.D. Cal. Feb. 25, 2010).

the foundational requirement that misrepresentation be “knowing,” not merely careless.²⁸³ Thus, in *Rossi* a takedown that could be verified to be improper — no file was ever in fact linked from a page stating movies were available — was held insufficient to support a 512(f) claim on the basis that the accusation was in good faith, though wrong.

E. Against Copyright Secondary Liability

The problems identified here are not due solely to the DMCA. The safe harbor amplifies features of the underlying copyright law — the risks to intermediary service providers of liability for the materials they host or reproduce for users — even as it offers service providers one way to mitigate those risks and provides copyright claimants a simple means of triggering secondary liability.

Much of the law and economics literature surrounding vicarious liability models a corporation or employer whose agent causes some tort harm. Where the direct-tortfeasor agent may be judgment-proof, the corporate principal is assigned liability to assure that the victim is compensated and to approach a socially optimal level of harm-prevention.²⁸⁴ Even here, scholars note that indirect liability is more expensive than direct liability because it includes both monitoring and precautionary costs.²⁸⁵

In the case of online copyright, by contrast, while the service-provider-intermediary is asked to assume the risks of vicarious liability, as a principal, its *functional* role is that of agent for end-user posters and speakers. A service provider (or several) is a necessary party to the end-user’s online communications, but it is the end-users’ interests that drive the communication and our policy concerns. This mismatch fuels concerns that secondary liability for copyright infringement over-deters speech.²⁸⁶ “Indirect liability has a significant

283. See 391 F.3d 1000, 1004–05 (9th Cir. 2004) (noting that “good faith” is a subjective standard and investigation to verify the accuracy of a DMCA claim is not required).

284. See, e.g., Lewis A. Kornhauser, *An Economic Analysis of the Choice between Enterprise and Personal Liability for Accidents*, 70 CALIF. L. REV. 1345 (1982); Giuseppe Dari Mattiacci & Francesco Parisi, *The Cost of Delegated Control: Vicarious Liability, Secondary Liability and Mandatory Insurance*, 23 INT’L REV. L. & ECON. 453 (2003); Alan O. Sykes, *The Economics of Vicarious Liability*, 93 YALE L.J. 1231 (1983); see also Mark A. Lemley & R. Anthony Reese, *Reducing Digital Copyright Infringement Without Restricting Innovation*, 56 STAN. L. REV. 1345, 1366 (2004) (“Vicarious liability in copyright law can be traced back to the doctrine of respondeat superior and was initially used to hold employers liable for infringements committed by their employees.”).

285. Mattiacci & Parisi, *supra* note 284, at 456.

286. See Lemley & Reese, *supra* note 284, at 1349–50 (“[W]hile courts can make decisions about direct infringement on a case-by-case basis, lawsuits based on indirect liability sweep together both socially beneficial and socially harmful uses of a program or service, either permitting both uses or condemning both.”); Alfred C. Yen, *Third-Party Copyright Liability After Grokster*, 91 MINN. L. REV. 184, 187 (2006) (“Third-party copyright liability

drawback, however, in that legal liability — even if carefully tailored — inevitably interferes with the legitimate use of implicated tools, services, and venues.”²⁸⁷

Finally, consider recent scholarship that looks at service providers as “platform” providers in two-sided markets.²⁸⁸ To the extent that service providers think of themselves as mediating between customers on multiple sides, and curating their platforms to maximize the profit and minimize hassle, we see further deviation from the neutral forum in which all, even the disagreeable, can speak. An Internet whose forums are all maintained by private, mostly-commercial actors is already far from a public square. Imposing liability risks on the forum hosts encloses it further.

V. REFORMING COPYRIGHT TAKEDOWN

As it now sits, the anchor for online speech is tenuous. Individual speakers lack security in the availability of hosting for their speech, and the public lacks assurance that it will be able to receive and maintain access to the full range of speech inputs to our ongoing conversations, be they scientific, literary, artistic, political, or merely fun. Copyright is emerging as the tool of choice for those who would disrupt online expression.

The uncertainty of underlying copyright law compounds the errors of the DMCA regime, pushing individuals toward self-censorship and their service providers to censorious takedown. The First Amendment requires us to correct these biases in substance and process. Following the recommendations of James Boyle,²⁸⁹ I argue that

benefits society by encouraging individuals to stop others from infringing, but those benefits come at a price because third-party defendants cannot focus precautions solely on infringers.”).

287. Douglas Lichtman & William Landes, *Indirect Liability for Copyright Infringement: An Economic Perspective*, 16 HARV. J.L. & TECH. 395, 409 (2003); see also Neal Kumar Katyal, *Criminal Law in Cyberspace*, 149 U. PA. L. REV. 1003, 1007–08 (2001).

288. See, e.g., Kevin Boudreau & Andrei Hagiu, *Platform Rules: Multi-Sided Platforms as Regulators* (SSRN Working Paper Series, Sept. 18, 2008), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1269966; Jörg Claussen, Tobias Kretschmer & Philip Mayrhofer, *Private Regulation by Platform Operators — Implications for Usage Intensity* (SSRN Working Paper Series, May 5, 2010), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1599458.

289. See James Boyle, *A Politics of Intellectual Property: Environmentalism for the Net?*, 47 DUKE L.J. 87, 111 (1997) (identifying a tension between “information” and “innovation” views of intellectual property, and decrying the current tendency to over-propertize and over-protect). In the context of environmentalism, Boyle noted that:

The environmental movement gained much of its persuasive power by pointing out that there were structural reasons that we were likely to make bad environmental decisions; a legal system based on a particular notion of what “private property” entailed, and an engineering or scientific system that treated the world as a simple, linearly related set of causes and effects. In both of these conceptual systems, the en-

copyright law must be assessed environmentally: what costs accompany the copyright incentive to creative expression?

Concentrating enforcement at service provider chokepoints, while the cheapest enforcement mechanism from a copyright-owner's perspective, imposes too much collateral cost on the speech environment. As it stands, copyright is not serving the cause of semiotic democracy or promoting human flourishing.²⁹⁰ Moreover, the chilling effect analysis indicates that over-deterrence is a problem deeper than the DMCA notice-and-takedown regime; it is a problem endemic to copyright law and its secondary liabilities. As copyright expands in scope, time, and breadth, its erroneous application and the chill of secondary liability assume greater significance.

Instead, we should calibrate and limit service provider liability to support free exchange of ideas. The most speech-hospitable, least biasing regime, I argue, is that of common carriage. Common carriage requires service providers to carry traffic on non-discriminatory terms, guaranteeing all equal access to transit or forum. The non-intellectual property regime of Section 230 of the Communications Decency Act ("Section 230") moves partially in that direction.²⁹¹ While it does not require service providers to carry all traffic, it does eliminate their risk of liability for user-posted speech, apart from intellectual property and criminal claims.

Section 230 protects the providers of "interactive computer services" from most liability for the speech of their users.²⁹² To achieve the statutory purposes of enabling lawful speech by reducing disincentives on service providers, courts have interpreted the provision broadly.²⁹³ As the Fourth Circuit put it in the early *Zeran* case, "law-

vironment actually *disappeared*; there was no place for it in the analysis.

Id.

290. See WILLIAM W. FISHER III, PROMISES TO KEEP 247 (2004); William Fisher, *Theories of Intellectual Property*, in NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY 168, 188–89 (Stephen R. Munzer ed., 2001); William W. Fisher III, *Property and Contract on the Internet*, 73 CHI.-KENT L. REV. 1203, 1217 (1998) ("In an attractive society, all persons would be able to participate in the process of meaning-making. Instead of being merely passive consumers of cultural artifacts produced by others, they would be producers, helping to shape the world of ideas and symbols in which they live.").

291. 47 U.S.C. § 230 (2006). This statutory provision was enacted to address contradictory and speech-constricting rulings regarding liability for online defamation. See Zittrain, *supra* note 61, at 262.

292. *Id.* § 230(c)(1).

293. Overall, Congress passed the Communications Decency Act to encourage service providers to reduce online access to indecent material or content deemed "harmful to minors." Section 230 was intended to remove the disincentive to monitor and moderate user-generated content that arose from notice-based liability. See David Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 409–11 (2010). The Supreme Court struck down the bulk of the Communications Decency Act, but left section 230 in place. See *ACLU v. Reno*, 521 U.S. 844 (1997).

suits seeking to hold a service provider liable for its exercise of a publisher's traditional editorial functions — such as deciding whether to publish, withdraw, postpone or alter content — are barred.”²⁹⁴ Section 230 thus bars any non-intellectual property theory of liability that seeks to hold the host liable as speaker, publisher, or distributor of user-posted content. However, Section 230 specifically excludes intellectual property and criminal claims from its protections.²⁹⁵

Outside the realm of intellectual property, the Section 230 shield removes many of the explicit legal pressures from service providers to remove content. They can set their own terms of service — choosing to maintain “family-friendly” environments, attempting to build communities, or taking a hands-off, anything goes approach.

Common carriage would go a step further, mandating that service providers take *all* traffic while behaving as conduits. Service providers would then be affirmatively discouraged from removing lawful speech.²⁹⁶ Common carriage is mandated in telecommunications for Title II “telecommunication services” — the telephone carriers are not liable for anything you might say or sing over the telephone, and forbidden from interfering with it.²⁹⁷ Focus has been moving steadily away from common carriage, however, as the FCC has instead classified all Internet access services as “information service,” with lesser access requirements than those imposed on common carriers.²⁹⁸ The FCC’s May 2010 “third way” proposal to “recognize the transmission component of broadband access service — and only this component — as a telecommunications service,” would at least provide a substrate on which more neutral hosting services could be anchored.²⁹⁹

If common carriage for service providers is unlikely to be realized, we could still take service providers out of the loop for user-driven copyright infringement with brighter lines of protection.

294. *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 330 (4th Cir. 1997).

295. *See* 47 U.S.C. § 230 (e)(1)–(2).

296. Some argue that common carriage rules would themselves impinge on the free speech of service providers. *See* Angela J. Campbell, *Publish or Carriage: Approaches to Analyzing the First Amendment Rights of Telephone Companies*, 70 N.C. L. Rev. 1071, 1132 (1992) (relating assertions by Bell Companies that “regulated utilities have First Amendment rights just like other citizens”) If common carriage prohibited service providers from speaking or from hosting communities with more structured terms, that would be problematic, but common carriage is only a base layer. So long as that layer remains neutral, both service providers and their users at all subsequent levels are free to speak and to set terms.

297. *See* JONATHAN E. NUECHTERLEIN & PHILIP J. WEISER, *DIGITAL CROSSROADS: AMERICAN TELECOMMUNICATIONS POLICY IN THE INTERNET AGE* 23 (2005).

298. *Id.* at 165–68.

299. *See* JULIUS GENACHOWSKI, FCC, *THE THIRD WAY: A NARROWLY TAILORED BROADBAND FRAMEWORK* (May 6, 2010), *available at* <http://www.broadband.gov/the-third-way-narrowly-tailored-broadband-framework-chairman-julius-genachowski.html>.

Modifications to the DMCA could minimize the risks of error by confining the takedown remedy to the most easily identifiable and verifiable cases of infringement. This limitation could be achieved by narrowing the class of uses for which takedown was available, stiffening the identification requirements, and better balancing the burdens of claim and response.

Thus, we might release service providers from any liability where the claimed infringement was less than entire commercial appropriation of a copyrighted work.³⁰⁰ If the duty to respond arose only on receipt of a notice that fully identified the claimed infringing work and pointed to a situation almost certain to be infringement, the service provider could cheaply compare the two, verify the complaint, and run a substantially smaller risk of erroneous takedown. Limiting takedowns to claimed commercial appropriation of entire works and requiring proof to be submitted along with the notification would enable service providers to make informed determinations and lessen the opportunities for abusive claims.

Substantial alterations to the structure of the DMCA would be necessary to correct the fundamental flaw that targets of notifications are presumed guilty, and punished with the loss of speech, before they can contest the charges. The focus of copyright law should be put back on the direct infringer, with claims redressed through damages rather than prior restraint. Even changes to the timing could help. Rather than “expeditious” takedown, content removal should be deferred until the poster has been notified and given an opportunity to respond.³⁰¹ Counter-notification would toll the takedown obligation immediately, eliminating the ten to fourteen business day downtime. Trimming the counter-notification requirements to match the minimal elements required for initial notice and eliminating the ten day holding period would help those who face erroneous takedown to recover quickly. Even if the counter-notification rate increased tenfold, it would be unlikely to come from the wholesale copyists and would still represent a tiny number compared to the takedowns.

Better balancing § 512(f) would entail a more minor fix. Currently, the sender of a takedown notice need only make a “good faith” declaration of infringement that is not “knowingly materially misleading.”³⁰² He swears under penalty of perjury only that he acts on authority of a copyright owner.³⁰³ Thus, so long as copyright holders don’t send their bots out *intending* to err, their failure to validate the

300. While this would still fail under Rebecca Tushnet’s argument that even entire copying may be protected speech, *see supra* note 87, it would relieve the pressures on fair use and greatly alleviate the burdens on speech short of full copying.

301. Urban and Quilter make a similar recommendation. *See* Urban & Quilter, *supra* note 160, at 688.

302. 17 U.S.C. § 512(f) (2006).

303. *Id.* § 512(c)(3)(A).

results of the scan or to check fair use defenses may be excused so long as it was in “good faith.” The law should require greater diligence: declarations on penalty of perjury to match those required by the respondent, and perhaps even a bond against erroneous claims. If a poster can prove speech was wrongly removed, she should not have to engage in protracted litigation — the *Lenz* case has been running since the June 4, 2007 takedown of Lenz’s video.³⁰⁴ Strengthening the counter-suit provisions could encourage a plaintiffs’ bar to take up these cases as private attorneys general. Stiffening the penalties against claimants who obtained takedowns through misrepresentation of infringement would encourage claimants to verify and support their claims of infringement or penalize them for failure to do so rather than allowing them to shift that burden to service providers and posters.

While the First Amendment information environment would be better served by reining in the copyright excesses of the DMCA and intermediary liability, the trend of policy is, regrettably, in the opposite direction.

Lessons from the errors and incentive problems surrounding copyright takedowns are particularly timely amid debate on the Anti-Counterfeiting Trade Agreement (“ACTA”) and so-called “three strikes” proposals requiring service providers to disconnect allegedly infringing customers after repeated warnings or infringements.³⁰⁵ Here too, copyright enforcement is put into private hands as injunctive relief, with even more serious impact on expression. Under these “graduated response” plans, service providers are required to impose on Internet users a series of increasing sanctions in response to notifications claiming copyright infringement: these may include warnings, fines or suspensions of service, and finally termination of Internet service.³⁰⁶

304. See *Lenz v. Universal Music Corp.*, 572 F. Supp. 2d 1150, 1152 (N.D. Cal. 2008).

305. Although the negotiating parties assert that the ACTA documents and discussions are matters of national secrecy, some documents have leaked amid intense public pressure. Law Professor Michael Geist maintains excellent coverage and discussion of the issue. See *ACTA Posts*, MICHAEL GEIST, http://www.michaelgeist.ca/index.php?option=com_tags&task=view&tag=acta&Itemid=408 (last visited Dec. 21, 2010); see also Cecilia Kang, *Secret Internet Copyright Talks Raise Concerns*, POST TECH (Nov. 5, 2009, 7:15 PM), http://voices.washingtonpost.com/posttech/2009/11/secret_internet_copyright_talk.html.

“Three strikes” provisions that have been proposed or enacted in national law include France’s HADOPI (passed, struck down by the constitutional court, and re-passed), New Zealand’s Copyright (Infringing File Sharing) Amendment Bill (dropped), and the UK’s Digital Economy Bill (as of March 20, passed the House of Lords). See Annemarie Bridy, *ACTA and the Specter of Graduated Response* (Am. Univ. Washington Coll. of Law, Program on Info. Justice & Intellectual Prop. Research Paper, 2010), available at <http://digitalcommons.wcl.american.edu/research/2/>.

306. See Bridy, *supra* note 305; Peter K. Yu, *The Graduated Response*, 62 FLA. L. REV. 1373, 1379–80 (2010).

While it had been on entertainment company agendas for some time before, graduated response was first enacted in the French law on the distribution and protection of creative works on the Internet (Loi Favorisant la Diffusion et la Protection de la Création sur Internet, HADOPI).³⁰⁷ Initially, the French Constitutional Court blocked enforcement of the law, finding that Internet accounts could be suspended only upon approval by a judge.³⁰⁸ The bill was revised to require that a judge issue the suspensions, rather than the same HADOPI agency that sends warning letters. After two warnings, Internet users could face suspension of Internet access up to a year long.³⁰⁹

The UK Digital Economy Bill,³¹⁰ passed in a “wash-up” just before the change of Parliament in March 2010, gives copyright owners a notification process similar to that of the U.S. DMCA “if it *appears* to a copyright owner that a subscriber to an internet access service has infringed the owner’s copyright by means of the service” or has allowed another to use the service to infringe.³¹¹ The Digital Economy Bill draft differs from the DMCA in recommending an appeals process with independent oversight, although this occurs only if a subscriber complains.³¹² The Secretary of State is given significant authority to rewrite the law, once passed; for example, “[t]he Secretary of State may at any time by order impose a technical obligation on [service providers].”³¹³ A wide-ranging group of public interest and political participants have expressed opposition.³¹⁴

307. See Nicolas Jondet, 38th Annual TPRC Conference: The French Copyright Authority (HADOPI), the Graduated Response and the Disconnection of Illegal File-Sharers (Oct. 2010).

308. See Eric Pfanner, *France Approves Wide Crackdown on Net Piracy*, N.Y. TIMES (Oct. 22, 2009), http://www.nytimes.com/2009/10/23/technology/23net.html?_r=1.

309. See CODE DE LA PROPRIÉTÉ INTELLECTUELLE art. 335-7, *available at* <http://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006069414&idArticle=LEGIARTI000006279194&dateTexte=&categorieLien=cid> (“Lorsque l’infraction est commise au moyen d’un service de communication au public en ligne, les personnes coupables des infractions prévues aux articles L. 335-2, L. 335-3 et L. 335-4 peuvent en outre être condamnées à la peine complémentaire de suspension de l’accès à un service de communication au public en ligne pour une durée maximale d’un an, assortie de l’interdiction de souscrire pendant la même période un autre contrat portant sur un service de même nature auprès de tout opérateur.”).

310. Digital Economy Bill, 2009–10, H.L. Bill [89], *available at* <http://www.publications.parliament.uk/pa/ld200910/ldbills/001/10001.6-12.html#j161>.

311. *Id.* cl. 124A(1) (emphasis added).

312. *Id.* cl. 124E(4).

313. *Id.* cl. 124H.

314. See Bobbie Johnson, *Rush To Pass Digital Bill Will ‘Sidestep Democracy,’* THE GUARDIAN (London), Mar. 19, 2010, <http://www.guardian.co.uk/technology/2010/mar/19/digital-bill-open-letter>; *Open Letter: Wash-up Not Appropriate for Controversial Disconnection Proposals*, THE GUARDIAN (London), Mar. 19, 2010, <http://www.guardian.co.uk/technology/2010/mar/19/digital-britain-file-sharing>.

In proposals leaked from the secret negotiations around ACTA, it is clear that intermediary liability — and not safe harbors from it — is critical to the U.S. negotiations. A leaked EU memorandum states:

“On the limitations from 3rd party liability: to benefit from safe-harbours, service providers need to put in place policies to deter unauthorised storage and transmission of IP infringing content (ex: clauses in customers’ contracts allowing, *inter alia*, a graduated response). . . . This Section 3 should also contain ‘broad’ provisions regarding notice-and-takedown mechanisms.”³¹⁵

A recent draft provides for safe harbors with similar exceptions and takedown conditions as the U.S. DMCA. The proposition that service providers are responsible for their users’ behavior or best situated to stop copyright infringements takes little account of the speech-chilling effect such enforcement power has.

The danger in these proposals is that intermediary liability or its notice-driven threat produces an information space skewed toward the commercial, popular, and bland. Instead of an open field for creative expression, political discourse, and dissent, intermediary liability will tend to constrain our choices as speakers and listeners. The space will favor the commercial speakers who can pay the service providers’ extra costs of responding to complaints or indemnify them against future demands.

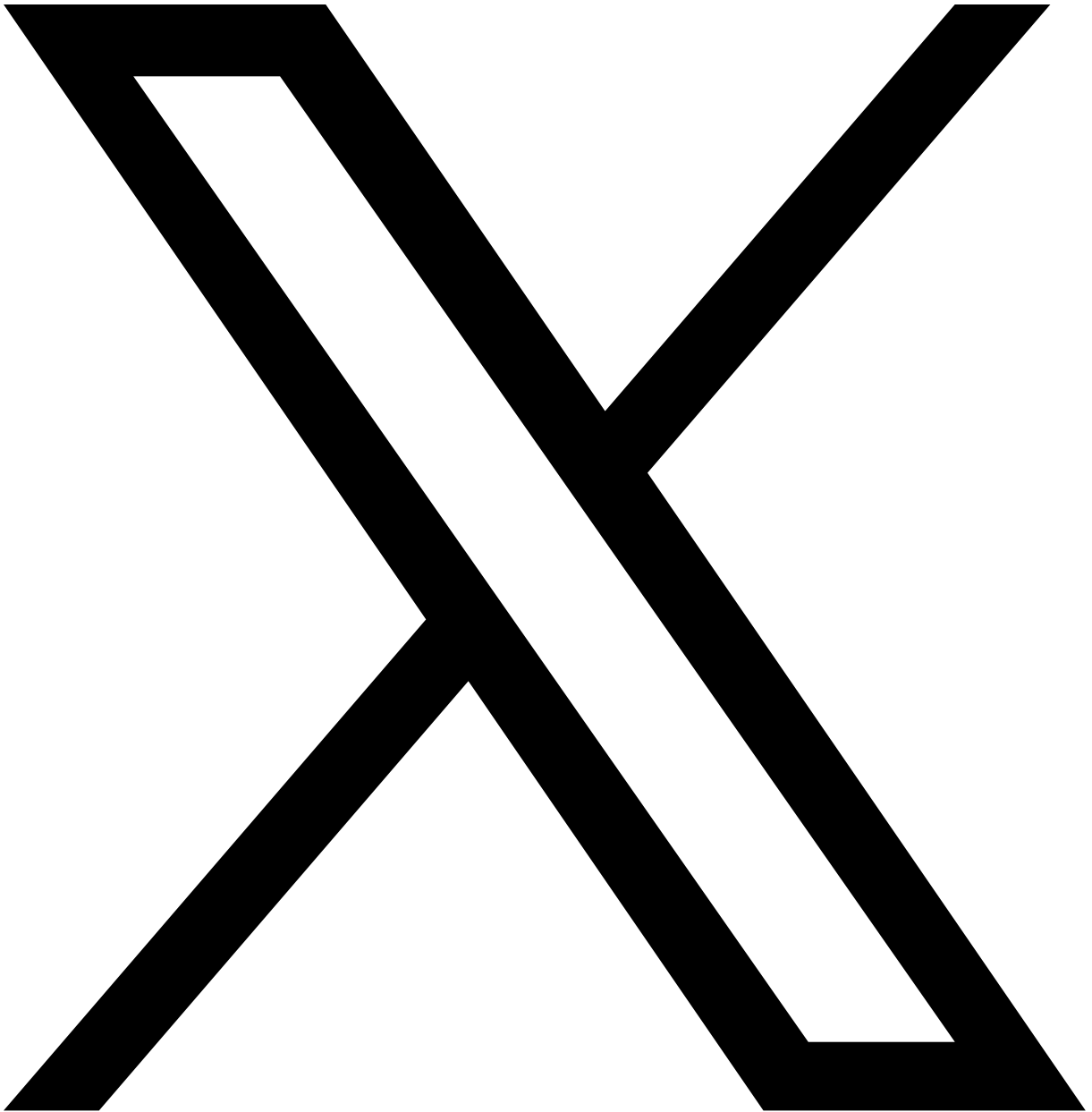
Popular mass content will find mirrors among fans who can keep it in circulation even as early posts are forced offline. Bland speech will face fewer challengers threatening to raise service providers’ hosting costs. Meanwhile speech that is non-commercial, minority, and challenging lacks these advantages. It will be vulnerable to takedown upon threat, and it will find fewer supporters willing to republish it. If its critical, parodic, or its opposition nature causes some to file DMCA takedown notices, even if unwarranted, this may be enough to bump the content offline and out of the public discourse. These errors in copyright’s author-protective mechanisms, eroding the very purpose of the copyright law and the First Amendment, should send us back to look for better-tailored enforcement measures.³¹⁶

315. ACTA — Internet Chapter (EC) No. 588/09 of 30 Sept. 2009, available at http://www.michaelgeist.ca/component/option,com_docman/task,doc_download/gid,26/.

316. See Lichtman & Landes, *supra* note 287, at 410. (“The core insight is that every mechanism for rewarding authors inevitably introduces some form of inefficiency, and thus the only way to determine the proper scope for indirect liability is to weigh its costs and benefits against the costs and benefits associated with other plausible mechanisms for rewarding authors.”).

Exhibit 16

[Skip to main content](#)



[Help Center \(https://help.x.com/en\)](https://help.x.com/en)

- [Using X](https://help.x.com/en/using-x) (https://help.x.com/en/using-x)
- [Managing your account](https://help.x.com/en/managing-your-account) (https://help.x.com/en/managing-your-account)
- [Safety and security](https://help.x.com/en/safety-and-security) (https://help.x.com/en/safety-and-security)
- [Rules and policies](https://help.x.com/en/rules-and-policies) (https://help.x.com/en/rules-and-policies)
- Resources ▼

 - [New user FAQ](https://help.x.com/en/resources/new-user-faq) (https://help.x.com/en/resources/new-user-faq)
 - [Glossary](https://help.x.com/en/resources/glossary) (https://help.x.com/en/resources/glossary)
 - [A safer X](https://help.x.com/en/resources/a-safer-twitter) (https://help.x.com/en/resources/a-safer-twitter)
 - [Accessibility](https://help.x.com/en/resources/accessibility) (https://help.x.com/en/resources/accessibility)

 - [Our rules](https://help.x.com/en/resources/rules) (https://help.x.com/en/resources/rules)
 - [My privacy](https://help.x.com/en/resources/how-you-can-control-your-privacy) (https://help.x.com/en/resources/how-you-can-control-your-privacy)
 - [How we address misinformation on X](https://communitynotes.twitter.com/guide/en/about/introduction)
(https://communitynotes.twitter.com/guide/en/about/introduction)
 - [Recommender Systems](https://help.x.com/en/resources/recommender-systems) (https://help.x.com/en/resources/recommender-systems)



[Contact Us](https://help.x.com/forms.html) (https://help.x.com/forms.html)

1. [Help Center](https://help.x.com/en) (https://help.x.com/en)
^
2. [Platform integrity and authenticity](https://help.x.com/en/rules-and-policies#platform-integrity-and-authenticity) (https://help.x.com/en/rules-and-policies#platform-integrity-and-authenticity).
^
3. Synthetic and manipulated media policy

Synthetic and manipulated media policy



1. [Help Center](https://help.x.com/en) ^ (https://help.x.com/en)
2. [Platform integrity and authenticity](https://help.x.com/en/rules-and-policies#platform-integrity-and-authenticity) ^ (https://help.x.com/en/rules-and-policies#platform-integrity-and-authenticity).

Synthetic and manipulated media policy

Overview

April 2023

You may not share synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm (“misleading media”). In addition, we may label posts containing misleading media to help people understand their authenticity and to provide additional context.

What is in violation of this policy

In order for content with **misleading media** (including images, videos, audios, gifs, and URLs hosting relevant content) to be labeled or removed under this policy, it must:

- Include media that is significantly and deceptively altered, manipulated, or fabricated, or
- Include media that is shared in a deceptive manner or with false context, and
- Include media likely to result in widespread confusion on public issues, impact public safety, or cause serious harm

We use the following criteria as we consider posts and media for labeling or removal under this policy as part of our ongoing work to enforce our rules and ensure healthy and safe conversations on X:

1. Is the content significantly and deceptively altered, manipulated, or fabricated?

In order for content to be labeled or removed under this policy, we must have reason to believe that media are significantly and deceptively altered, manipulated, or fabricated. Synthetic and manipulated media take many different forms and people can employ a wide range of technologies to produce these media. Some of the factors we consider include:

- whether media have been substantially edited or post-processed in a manner that fundamentally alters their composition, sequence, timing, or framing and distorts their meaning;
- whether there are any visual or auditory information (such as new video frames, overdubbed audio, or modified subtitles) that has been added, edited, or removed that fundamentally changes the understanding, meaning, or context of the media;
- whether media have been created, edited, or post-processed with enhancements or use of filters that fundamentally changes the understanding, meaning, or context of the content; and
- whether media depicting a real person have been fabricated or simulated, especially through use of artificial intelligence algorithms

We will not take action to label or remove media that have been edited in ways that do not fundamentally alter their meaning, such as retouched photos or color-corrected videos.

In order to determine if media have been significantly and deceptively altered or fabricated, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been altered or fabricated, we may not take action to label or remove them.

2. Is the content shared in a deceptive manner or with false context?

We also consider whether the context in which media are shared could result in confusion or suggests a deliberate intent to deceive people about the nature or origin of the content, for example, by falsely claiming that it depicts reality. We assess the context provided alongside media to see whether it provides true and factual information. Some of the types of context we assess in order to make this determination include:

- whether inauthentic, fictional, or produced media are presented or being endorsed as fact or reality, including produced or staged works, reenactments, or exhibitions portrayed as actual events;
- whether media are presented with false or misleading context surrounding the source, location, time, or authenticity of the media;
- whether media are presented with false or misleading context surrounding the identity of the individuals or entities visually depicted in the media;
- whether media are presented with misstatements or misquotations of what is being said or presented with fabricated claims of fact of what is being depicted

We will not take action to label or remove media that have been shared with commentary or opinions that do not advance or present a misleading claim on the context of the media such as those listed above.

In order to determine if media have been shared in a deceptive manner or with false context, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been shared with false context, we will not label or remove the content.

3. Is the content likely to result in widespread confusion on public issues, impact public safety, or cause serious harm?

Posts that share misleading media are subject to removal under this policy if they are likely to cause serious harm. Some specific harms we consider include:

- Threats to physical safety of a person or group
- Incitement of abusive behavior to a person or group
- Risk of mass violence or widespread civil unrest
- Risk of impeding or complicating provision of public services, protection efforts, or emergency response
- Threats to the privacy or to the ability of a person or group to freely express themselves or participate in civic events, such as:
 - Stalking or unwanted and obsessive attention
 - Targeted content that aims to harass, intimidate, or silence someone else's voice
 - Voter suppression or intimidation

We also consider the time frame within which the content may be likely to impact public safety or cause serious harm, and are more likely to remove content under this policy if immediate harm is likely to result.

Posts with misleading media that are not likely to result in immediate harm but still have a potential to impact public safety, result in harm, or cause widespread confusion towards a public issue (health, environment, safety, human rights and equality, immigration, and social and political stability) may be labeled to reduce their spread and to provide additional context.

While we have other rules also intended to address these forms of harm, including our policies on violent threats, civic integrity, and hateful conduct, we will err toward removal in borderline cases that might otherwise not violate existing rules for Posts that include misleading media.

What is not a violation of this policy

We seek to protect public conversation surrounding various issues. Media often accompany these conversations and encourage further discourse. In the absence of other policy violations, the following are generally not in violation of this policy:

- **Memes or satire**, provided these do not cause significant confusion about the authenticity of the media;
- **Animations, illustrations, and cartoons**, provided these do not cause significant confusion about the authenticity of the media.
- **Commentary, reviews, opinions, and/or reactions**. Sharing media with edits that only add commentary, reviews, opinions, or reactions allows for further debate and discourse relating to various issues and are not in violation of this policy.
- **Counterspeech**. We allow for direct responses to misleading information which seek to undermine its impact by correcting the record, amplifying credible information, and educating the wider community about the prevalence and dynamics of misleading information.

What happens if you violate this policy?

The consequences for violating our synthetic and manipulated media policy depends on the severity of the violation.

Post Deletion

For high-severity violations of the policy, including misleading media that have a serious risk of harm to individuals or communities, we will require you to remove this content.

Labeling

In circumstances where we do not remove content which violates this policy, we may provide additional context on posts sharing the misleading media where they appear on X. This means we may:

- Apply a label and/or warning message to the post
- Show a warning to people before they share or like the post;
- Reduce the visibility of the post on the platform and/or prevent it from being recommended;
- Turn off likes, replies, and Reposts; and/or
- Provide a link to additional explanations or clarifications, such as relevant X policies.

In most cases, we will take a combination of the above actions on posts we label.

Account locks

If we determine that an account has advanced or continuously shares harmful misleading narratives that violate the synthetic and manipulated media policy, we may temporarily reduce the visibility of the account or lock or suspend the account.

If you believe that your account was locked or suspended in error, you can [submit an appeal](https://help.twitter.com/forms/general?subtopic=suspended) (<https://help.twitter.com/forms/general?subtopic=suspended>).

Additional resources

Learn more about [our range of enforcement options](#).

Share this article



© 2024 X Corp.

[Cookies](https://help.x.com/rules-and-policies/twitter-cookies) (<https://help.x.com/rules-and-policies/twitter-cookies>)

[Privacy](https://x.com/privacy) (<https://x.com/privacy>)

[Terms and conditions](https://x.com/tos) (<https://x.com/tos>)

English



[Help Center](https://help.x.com/en) (<https://help.x.com/en>)

- [English](https://help.x.com/en/rules-and-policies/manipulated-media) (https://help.x.com/en/rules-and-policies/manipulated-media)
- [Español](https://help.x.com/es/rules-and-policies/manipulated-media) (https://help.x.com/es/rules-and-policies/manipulated-media)
- [日本語](https://help.x.com/ja/rules-and-policies/manipulated-media) (https://help.x.com/ja/rules-and-policies/manipulated-media)
- [한국어](https://help.x.com/ko/rules-and-policies/manipulated-media) (https://help.x.com/ko/rules-and-policies/manipulated-media)
- [Português](https://help.x.com/pt/rules-and-policies/manipulated-media) (https://help.x.com/pt/rules-and-policies/manipulated-media)
- [Deutsch](https://help.x.com/de/rules-and-policies/manipulated-media) (https://help.x.com/de/rules-and-policies/manipulated-media)
- [Türkçe](https://help.x.com/tr/rules-and-policies/manipulated-media) (https://help.x.com/tr/rules-and-policies/manipulated-media)
- [Français](https://help.x.com/fr/rules-and-policies/manipulated-media) (https://help.x.com/fr/rules-and-policies/manipulated-media)
- [Italiano](https://help.x.com/it/rules-and-policies/manipulated-media) (https://help.x.com/it/rules-and-policies/manipulated-media)
- [العربية](https://help.x.com/ar/rules-and-policies/manipulated-media) (https://help.x.com/ar/rules-and-policies/manipulated-media)
- [Nederlands](https://help.x.com/nl/rules-and-policies/manipulated-media) (https://help.x.com/nl/rules-and-policies/manipulated-media)
- [Bahasa Indonesia](https://help.x.com/id/rules-and-policies/manipulated-media) (https://help.x.com/id/rules-and-policies/manipulated-media)
- [Русский](https://help.x.com/ru/rules-and-policies/manipulated-media) (https://help.x.com/ru/rules-and-policies/manipulated-media)
- [हिंदी](https://help.x.com/hi) (https://help.x.com/hi)
- [தமிழ்](https://help.x.com/ta) (https://help.x.com/ta)
- [עברית](https://help.x.com/he) (https://help.x.com/he)
- [简体中文](https://help.x.com/zh-cn) (https://help.x.com/zh-cn)
- [繁體中文](https://help.x.com/zh-tw) (https://help.x.com/zh-tw)
- [ภาษาไทย](https://help.x.com/th) (https://help.x.com/th)
- [Tiếng Việt](https://help.x.com/vi) (https://help.x.com/vi)
- [Melayu](https://help.x.com/ms) (https://help.x.com/ms)
- [ইংরেজি](https://help.x.com/bn) (https://help.x.com/bn)
- [Filipino](https://help.x.com/fil) (https://help.x.com/fil)
- [فارسی](https://help.x.com/fa) (https://help.x.com/fa)
- [Dansk](https://help.x.com/da) (https://help.x.com/da)
- [Suomi](https://help.x.com/fi) (https://help.x.com/fi)
- [Svenska](https://help.x.com/sv) (https://help.x.com/sv)
- [Norsk](https://help.x.com/no) (https://help.x.com/no)
- [Polski](https://help.x.com/pl) (https://help.x.com/pl)
- [Magyar](https://help.x.com/hu) (https://help.x.com/hu)
- [Română](https://help.x.com/ro) (https://help.x.com/ro)
- [Українська](https://help.x.com/uk) (https://help.x.com/uk)
- [मराठी](https://help.x.com/mr) (https://help.x.com/mr)
- [ગુજરાતી](https://help.x.com/gu) (https://help.x.com/gu)
- [Български](https://help.x.com/bg) (https://help.x.com/bg)
- [Català](https://help.x.com/ca) (https://help.x.com/ca)
- [Hrvatski](https://help.x.com/hr) (https://help.x.com/hr)
- [Српски](https://help.x.com/sr) (https://help.x.com/sr)
- [Slovenčina](https://help.x.com/sk) (https://help.x.com/sk)
- [ಕನ್ನಡ](https://help.x.com/kn) (https://help.x.com/kn)

- پښتو ژبې (<https://help.x.com/ps>)
- Dari (<https://help.x.com/fa-af>)
- Oromo (<https://help.x.com/om>)
- Tigrinya (<https://help.x.com/ti>)
- English (<https://help.x.com/ckb>)
- Lietuvių (<https://help.x.com/lt>)
- Latviešu (<https://help.x.com/lv>)
- Malti (<https://help.x.com/mt>)
- Slovenščina (<https://help.x.com/sl>)
- Gaeilge (<https://help.x.com/ga>)
- Lus Hmoob (<https://help.x.com/hmn>)
- Հայերեն (<https://help.x.com/hy>)
- ខ្មែរ (<https://help.x.com/km>)

Exhibit 17

Search help articles

How to identify AI content on Meta products

Updated: 6 weeks ago

Labeling and identifying AI content is different for ads.

Generative artificial intelligence (generative AI) is a tool that helps create or modify content such as images and text in new ways.

How generative AI works

Generative AI models are trained using billions of pieces of information. To generate new content, generative AI models learn about patterns and relationships between these pieces of information based on prompts or instructions.

For example, you might type a request in a generative AI tool asking for an image of 3 brown dogs running through a meadow of flowers. Or you might request a joke written in the style of a historical figure, like Shakespeare.

How AI-generated content is identified on Meta products

Content created or modified using AI tools may be identified and labeled to help promote transparency across Meta products.

There are many AI tools that can create or modify content. Some content that is made or modified using one of these tools may have an industry-standard signal that identifies the content as being created or modified using AI tools. For example, a person might edit an image using a common photo editing tool to share with their followers.

If a photo editing tool uses AI to change the size or color of an image, then the image may have a signal that shows it was edited using AI. These signals are read by Meta's systems to

When AI-generated content is labeled on Meta products

There are a few ways organic content, or [content that is not an ad](#), will be labeled on Meta products.

- **When Meta's systems detect AI signals in organic content**

Content made with third-party tools may contain signals that show whether the content was created or modified with AI. Content with signals indicating that it was created with AI will be labeled **AI info**. Content with signals indicating that it was modified using AI will not have the **AI info** label directly on the post, story, reel or thread. You can learn more about how the post, story, reel or thread was modified by AI by clicking the **...** menu.

- **When people label their AI-generated content**

People using Meta products are also able, and in some cases required, to label their content with **AI info** when they share AI-generated content or content that was modified using AI.

Meta may continue to update the approach around transparency as more is learned about AI-generated content across Meta technologies.

Other AI labels you might see

Any content that is created or edited using Meta's AI tools and shared to Facebook, Instagram, or Threads as a post, story, reel, or thread may be automatically labeled as AI content or in some cases, feature a visible watermark.

Note: Not all AI content contains the information needed to identify it.

When the **AI info** label is required

Meta requires an AI label when content has photorealistic video or realistic-sounding audio that was digitally created, modified or altered, including with AI.

Note: There may be penalties for content shared without a label when it is required.

Meta does not require a label for images that have been created or modified with AI. However, these images may still receive a label if the systems detect that they were AI-generated or if they were modified using AI.

Examples of digitally created content that requires a label:

- A video that appears realistic of a group of people walking around an outdoor market
- An audio file of two people talking
- A song created using AI-generated vocals
- A reel narrated with a realistic AI-generated voiceover

Examples of digitally created content that does not require a label:

- Video of an outdoor landscape, created in a style resembling a cartoon
- Image of a person riding a bull (Meta does not require a label for images)

You can read more about this requirement in the [Transparency Center](#).

Note: We will update our transparency approach as we learn more about AI-generated content on our platforms. Learn more about generative AI in the [Privacy Center](#).

➔ [Share](#)

Was this article helpful?

Yes

No

 Meta



[Site terms and policies](#)

[Community standards](#)

Privacy policy

Terms

Cookie policy

Virtual reality



Smart glasses



Support and legal



Our actions



About us



Our community



App support



English (US)

META QUEST

Meta Quest: *Parents:* Important guidance & safety warnings for children's use [here](#). Using Meta Quest requires an account and is subject to requirements that include a minimum age of 10 (requirements may vary by country). See [meta.com/quest/terms](#) and the parent's info page at [meta.com/quest/parent-info](#). Certain apps, games and experiences may be suitable for a more mature audience. META QUEST FEATURES, FUNCTIONALITY, AND CONTENT NOTICE: Features, functionality and content are subject to change or withdrawal at any time, may not be available in all areas or languages or may be restricted; may require enabled software or service activation, and additional terms, conditions and/or charges may apply.

META QUEST IMPORTANT SAFETY NOTICE <https://www.meta.com/quest/quest-2-facial-interface-recall/>.

Financing Options. You may be offered financing options for your Meta purchases. Learn more [here](#).

***Based on the graphic performance of the Qualcomm Snapdragon XR2 Gen 2 vs XR2 Gen 1 on Meta Quest 2

Meta AI and voice commands only in select countries and languages. Please check local availability. Meta account and Meta View App required. For ages 13+ only. Requires compatible phone with Android or iOS operating system plus wireless internet access. Features, functionality and content are subject to change or withdrawal at any time. Additional account registration, terms and fees may apply. Software updates may be required. Performance may vary based on user location, device battery, temperature, internet connectivity and interference from other devices, plus other factors. User must comply with all applicable local laws and regulations, especially relating to privacy. May interfere with personal medical devices. Check manufacturer [Safety & Warranty Guide](#) and [FAQs](#) for more product information, including [battery life](#).

©2024 Meta.

Exhibit 18

Disclosing use of altered or synthetic content

We encourage creators' innovative and responsible use of content editing or generation tools. At the same time, we recognize that viewers want to know if what they're watching or listening to is real.

To help keep viewers informed about the content they're viewing, we require creators to disclose content that is meaningfully altered or synthetically generated when it seems realistic.

Creators must disclose content that:


- Makes a real person appear to say or do something they didn't do
- Alters footage of a real event or place
- Generates a realistic-looking scene that didn't actually occur

This could include content that is fully or partially altered or created using audio, video or image creation or editing tools.

Disclose using the 'altered content' setting in YouTube Studio

To disclose content that is meaningfully altered or synthetically generated, the 'altered content' setting is available to creators using YouTube Studio on a computer or mobile device. We'll expand this setting to other devices and YouTube apps in the future.

After a creator selects this field and uploads content, a label will appear in their [video's expanded description](#).

Creators who make a YouTube Short using [Dream Track](#) or [Dream Screen](#) , YouTube's generative artificial intelligence (AI) tools, don't need to take extra steps to disclose. The tool will automatically disclose the use of AI for creators. For other AI tools, creators need to disclose their use during the upload flow.

Examples of altered or synthetic content

The following list includes examples of altered or synthetic content. Altered or synthetic content can include content that is fully or partially altered or created using any audio, video, image creation or editing tools. Realistic content and meaningful changes require disclosure, while unrealistic or minor edits don't. Keep in mind, this isn't a complete list.

Doesn't require disclosure by creators	Does require disclosure by creators
<ul style="list-style-type: none"> • Applying beauty filters 	<ul style="list-style-type: none"> • Digitally generating or altering content to replace the face of one individual with another's
<ul style="list-style-type: none"> • Synthetically generating or extending a backdrop to simulate a moving car 	<ul style="list-style-type: none"> • Digitally altering a famous car chase scene to include a celebrity who wasn't in the original movie
<ul style="list-style-type: none"> • Using effects to enhance previously recorded audio 	<ul style="list-style-type: none"> • Simulating audio to make it sound as if a medical professional gave advice when the professional did not actually give that advice
<ul style="list-style-type: none"> • Using an AI-generated animation of a missile in a video 	<ul style="list-style-type: none"> • Showing a realistic depiction of a missile fired toward a real city

[Examples of content creators don't have to disclose](#)



Disclose altered or synthetic content

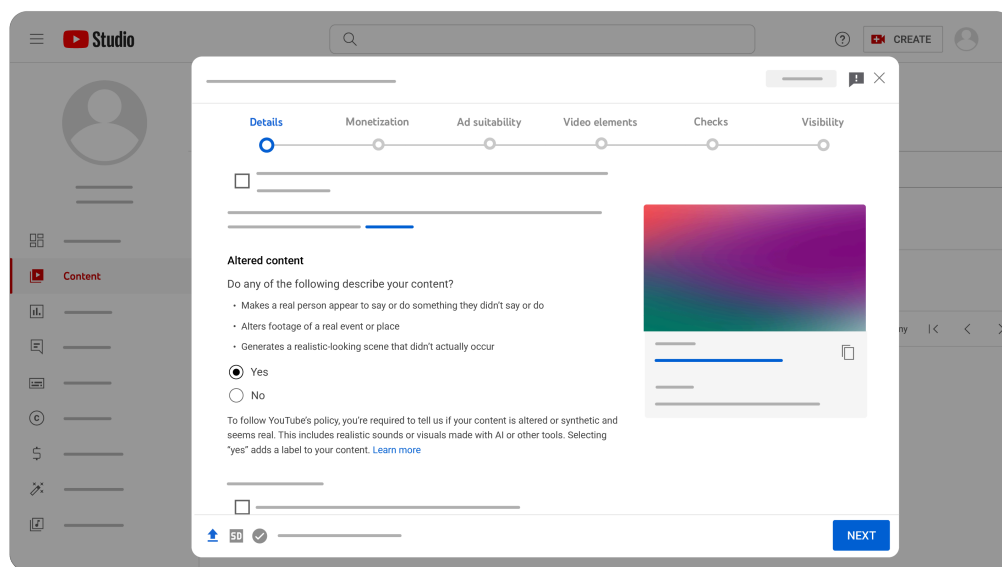
We require creators to disclose meaningfully altered or synthetically generated content that seems realistic. Creators can make this disclosure during the upload process.

[Computer](#) [Android](#) [iPhone & iPad](#)

1. Go to [YouTube Studio](#).
2. Follow the steps to upload content.
3. In the Details section, under “Altered content,” select **Yes** if your content meets the disclosure requirements.
4. Continue to select other video details.

If a creator makes a YouTube Short that uses one of YouTube’s generative artificial intelligence (AI) effects (such as YouTube’s [Dream Track](#) or [Dream Screen](#)), no additional steps are needed to disclose currently. The tool will automatically disclose the use of AI for creators.

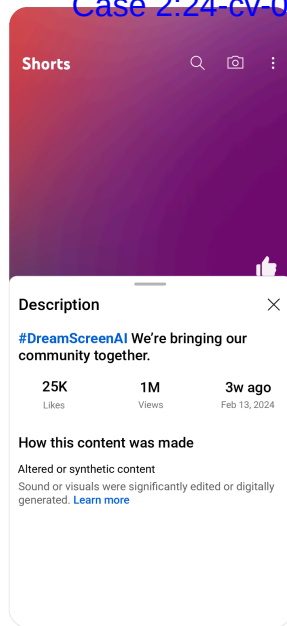
To help creators, we may proactively select disclosure on their behalf if they disclose the use of altered or synthetic content in the title or description of their video.



What happens after creators disclose

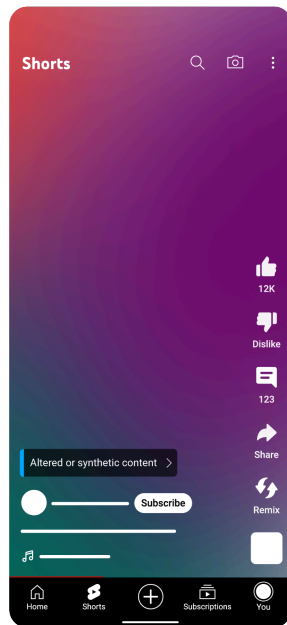
If creators select “Yes” to indicate that their content is altered or synthetic, we’ll add a label to their video’s **description field**. These labels will show for viewers watching YouTube videos on a mobile device or tablet for now.

Label in the expanded description field



Additional label for sensitive content

High quality and timely information about elections, ongoing conflicts, natural disasters, finance, or health is very important. This type of information can greatly affect the well-being, financial security, or safety of people and communities. For content about sensitive topics like these, a more prominent label in the **video player** may also appear for added transparency.



Other impacts of disclosing

Disclosing content as altered or synthetic won't limit a video's audience or impact its eligibility to earn money.

Risks of not disclosing

It can be misleading if viewers think a video is real, when it's actually been meaningfully altered or synthetically generated to seem realistic.

When content is undisclosed, in some cases YouTube may take action to reduce the risk of harm to viewers by proactively applying a label that creators will not have the option to remove. Additionally,

creators who consistently choose not to disclose this information may be subject to penalties from YouTube, including removal of content or suspension from the YouTube Partner Program.

Remember, we apply our [Community Guidelines](#) to all content on YouTube, regardless of whether it's altered or synthetic.

Need more help?
Try these next steps:



Post to the help community

Get answers from community members

Exhibit 19



Getting started

Using TikTok

Creating videos

Making a post

Camera tools

Creating effects on mobile

Effects

Sounds

Editing, posting, and deleting

TikTok Studio

Editing TikTok videos and photos

Duets

Stitch

TikTok Stories

TikTok Notes

Create playlists of your videos

Credit a video

Add Yours

Books

Movies and TV shows

Accessibility for your videos

About AI-generated content



Support nonprofits on TikTok

Exploring videos



For You

Friends Tab

STEM feed

TikTok Shop

Watching Stories on TikTok

Liking

Sharing on TikTok

Repost

Accessibility for watching videos

Your favorite books

Your favorite movies and TV shows

Add songs from TikTok

Watch videos in a Series

Watch videos in a playlist

How TikTok recommends content

Donations on TikTok

Cast TikTok to a TV

TikTok for Apple Vision Pro

Video and photo downloads

Discover and search

Messaging and notifications





Home

Mentions on TikTok

Streaks

Notifications

TikTok stickers

Followers and Following



Following and unfollowing

Finding friends from your contacts

Removing followers

Blocking someone

Finding your blocked list

Growing your audience



How to grow your audience

Creator Search Insights

Verified accounts on TikTok

Personal and Business Accounts on TikTok

Government, Politician, and Political Party Accounts

My posts aren't getting views

How can creators monetize on TikTok?

Use Promote to grow your TikTok audience

Report a problem



Share feedback

Report a problem

Account and privacy settings

TikTok LIVE, Gifts, and wallet



Monetize on TikTok



About AI-generated content

Jump to a section

[What is AI-generated content?](#) • [What are the requirements for posting AI-generated content to TikTok?](#) • [How is AI-generated content labeled on TikTok?](#) • [Why should creators label AI-generated content?](#) • [What kinds of AI-generated content are prohibited entirely on TikTok?](#) • [How to apply the AI-generated content label](#)

What is AI-generated content?

AI-generated content includes images, video and/or audio that is generated or modified by deep- or machine-learning processes. This content may include depictions of real people that may be highly realistic or created in a particular artistic style (e.g. painting, cartoons, and anime).

Examples of AI-generated content include the following:

- Video featuring a real person speaking, whose image, voice, and/or words are altered or modified by AI
- Video or image featuring a scene or event that occurred in the real world, but has been altered or modified by AI
- Entirely AI-generated videos or images of real or fictional people, places, and events

TikTok?

To support authentic and transparent experiences for our community, we encourage creators to label content that has been completely generated or significantly edited by AI. We consider content that's significantly edited by AI as that which uses real images/video as source material, but has been modified by AI beyond minor corrections or enhancements, including synthetic images/video in which:

- The primary subjects are portrayed doing something they didn't do, e.g. dancing;
- The primary subjects are portrayed saying something they didn't say, e.g. by AI voice cloning; or
- The appearance of the primary subject(s) has been substantially altered, such that the original subject(s) is no longer recognizable, e.g. with an AI face-swap.

We also require creators to label all AI-generated content where it contains realistic images, audio, and video, as explained in our [Community Guidelines](#).

How is AI-generated content labeled on TikTok?

Creators can disclose content as AI-generated directly on the video by adding text, a hashtag sticker, or context in the video's description.

We also provide several labels that can let viewers know when AI was used:

- **AI-generated** label: We may automatically apply this label to content that we detect was created or edited by AI. Please note that if you made your AI-generated content using only TikTok effects, then you don't need to label it as the effect name provides context to the viewer. If you independently altered your video with AI in addition to the TikTok effect you used, we still ask that you follow our labeling guidelines.
- Our **Creator labeled as AI-generated** label: Creators apply this label themselves to indicate that their content was completely generated or significantly edited by AI. Please note that misleadingly labeling unaltered content with this label is a violation of our [Terms of Service](#) and may result in the removal of content.

If you see a video on TikTok that you believe violates our [Synthetic and Manipulated Media](#) policy, please [report it to us](#).

Why should creators label AI-generated content?

We ask creators to disclose their AI-generated content in order to:

- Help prevent the spread of misleading information on TikTok by making clear to viewers which content is unaltered and which is altered or modified by AI technology.
- Follow our [Community Guidelines](#) on integrity and authenticity, including our [misinformation or impersonation policies](#). We may remove content that violates our Community Guidelines policies.

What kinds of AI-generated content are prohibited entirely on TikTok?

The following types of AI-generated content containing the likeness (visual or audio) of a real or fictional person aren't allowed, even if disclosed with the AI-generated content label, and may be removed:

- A public figure when used for political or commercial endorsements. We define public figures as adults (18 years and older) with a significant public role, such as a government official, politician, business leader, or celebrity.
- A private figure. We define private figures as any person who isn't a public figure, including people under 18 years old.

How to apply the AI-generated content label

To turn on the AI-generated content setting before you post on the TikTok app:

1. On the **Post** screen, tap the **More options** button.
2. Turn the **AI-generated content** setting on.
3. Tap **Post**.

Once posted, your video will be labeled as **Creator labeled as AI-generated** and can't be changed.

Was this helpful?

 Yes

 No

Helpful links

[Creating an account](#)

[Setting up your profile](#)

[Creating a TikTok video](#)



Company

[About TikTok](#)

[Newsroom](#)

[Contact](#)

[Careers](#)



Home

TikTok for Developers

Effect House

Advertise on TikTok

TikTok Rewards

TikTok Browse

TikTok Embeds

Resources

Help Center

Safety Center

Creator Portal

Community Guidelines

Transparency

Accessibility

Legal

Terms of Service

English



© 2024 TikTok

Exhibit 20



What can we help you with?

[Snapchat Support](#) > [Using Snapchat](#) > [Generative AI on Snapchat](#) >

Generative AI on Snapchat

Table of Contents:

- [What is generative AI?](#)
- [Which Snapchat features use generative AI?](#)
- [How can I recognize generative AI tools on Snapchat?](#)
- [What is the Snap Ghost with sparkles watermark?](#)
- [What are the dos and don'ts of generative AI on Snapchat?](#)
- [How is information I share with generative AI features used?](#)
- [What can I do if I don't like an image or response created with generative AI?](#)

What is generative AI?

Generative AI is a type of technology that learns from a large amount of data to create entirely new content — like text, images, and videos. Machine learning models, like generative AI, can get more accurate over time by recognizing patterns. For example, it can recognize that an image of a dog is a dog after seeing many pictures of different breeds by identifying similarities in images. These models evolve (for example, when they are trained with additional data) so that over time they get better, eventually recognizing more attributes (for example, a specific breed of dog).

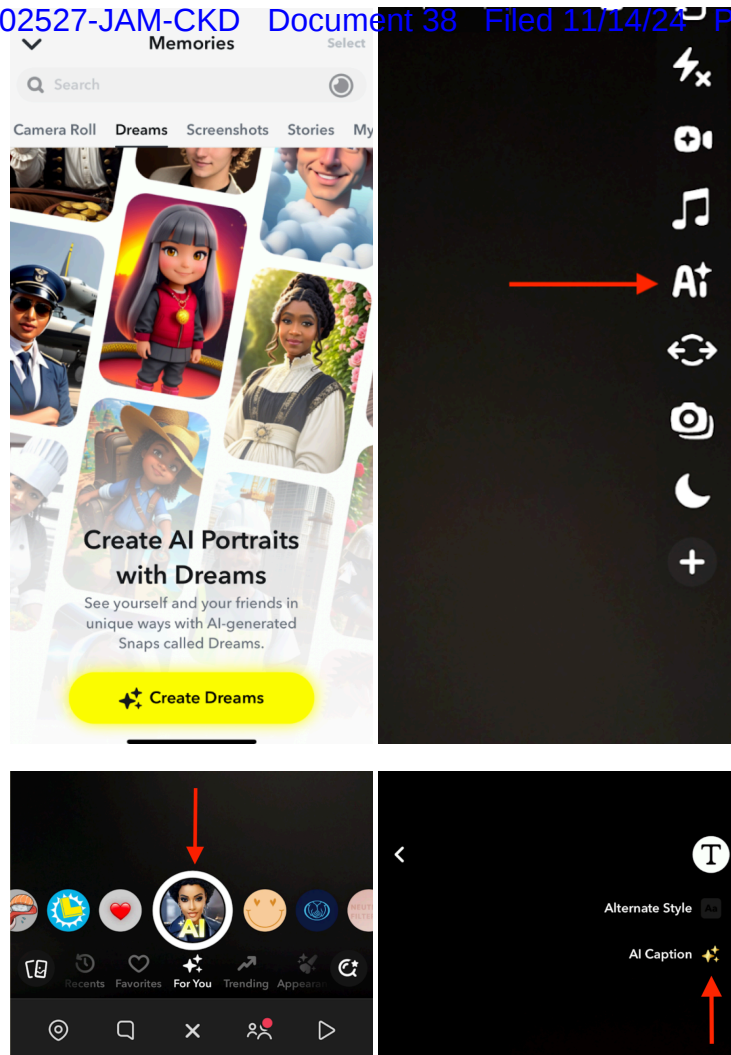
Which Snapchat features use generative AI?

We are constantly working on new ways to enhance your experience with generative AI. Many new features on Snapchat are powered by generative AI, like [AI Lenses](#), [My AI](#), [Dreams](#), and [AI Snaps](#).

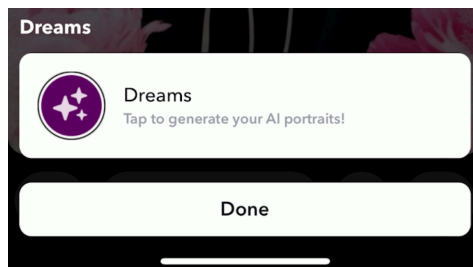
How can I recognize generative AI tools on Snapchat?

We may indicate that a feature in Snapchat is powered by generative AI in a number of ways, including using the sparkle icon ✨, specific disclaimers, [Context Cards](#), or tool tips. When you see these contextual symbols or other indicators in Snapchat, you should know that you are interacting with AI and not a human, or viewing content that has been produced using AI and does not depict real world scenarios.

Here are examples of what these AI indicators look like in Snapchat:



When some AI generated images are shared with you on Snapchat, we may include a Context Card to let you know that the image was created with a generative AI-powered feature.



What is the Snap Ghost with sparkles watermark?

Some generative AI-powered features, like [Dreams](#) and [AI Snaps](#), allow you to create or edit images. When you export or save a generated image to Camera Roll, a watermark of a Snap Ghost with sparkles may be added to those images. The purpose of these watermarks is to provide transparency that the image was created with generative AI and is not real or based on real events, even if it is a realistic style.

Here's an example of what the watermark looks like:



⚠ Please Note: Not all AI generated images will include a Context Card or watermark. Images created with non-Snap products may not be labeled as AI generated.

What are the dos and don'ts of generative AI on Snapchat?

Across all of Snapchat's tools and services we strive to empower people to express themselves, live in the moment, learn about the world, and have fun together. We are so excited by generative AI technology's potential to help our community continue to unlock their creativity and imagination.

Generative AI is a quickly evolving technology and we are committed to developing it responsibly. To make Snapchat's generative AI features safe and meaningful for all users, please adhere to these guidelines:

Dos:

- Get creative with generative AI-powered features to express yourself and learn new things!
- Use Snapchat responsibly, following our [Terms of Service](#) and [Community Guidelines](#), which, among other things, prohibit generating harmful, inappropriate or misleading content.

Don'ts:

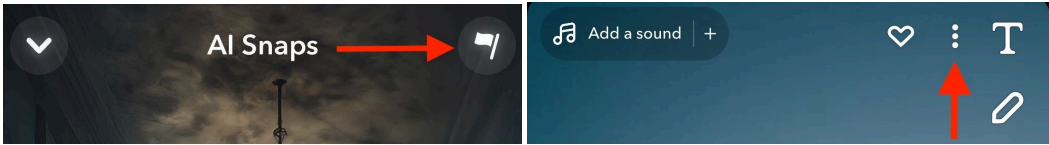
- Do not share sensitive or private information. For example, do not share health related information with My AI or images of people who have not given you permission to use their image or likeness in generative AI imagery features.
- Do not assume generative AI outputs are true or depict real events. Generative AI can and will make mistakes, and as a result outputs may be incorrect, inappropriate, or wrong.
- Do not expect generated images to reflect how you look in real life or accurately maintain vital image attributes. This is technology intended to alter or transform an image. For example, if you typically wear a headscarf, generated images may depict you without one.
- Do not remove Snap's Ghost with sparkles watermark as it is a violation of our Terms.

How is information I share with generative AI features used?

When you share text and images with generative features, the AI uses that information to provide the generated text or image or response to you. If you ask our AI tools to generate a dog riding a horse, that text prompt "dog riding a horse" is processed by the model and used to deliver that image to you. For additional information on how Snap uses your information, please see our [Privacy Policy](#) and [Privacy by Product](#) page.

What can I do if I don't like an image or response created with generative AI?

Many of Snapchat's generative AI features have reporting tools. To get help, the flag icon in the top right corner of your feed. Snap wants your feedback and we encourage you to report any content that violates our [Community Guidelines](#).



Was this article helpful?

☒ Yes

☐ No

Articles in this section

[Generative AI on Snapchat](#)

Related articles

- [How do AI Snaps work with Snapchat+?](#)
- [What is Snapchat+?](#)
- [About Context Cards on Snapchat](#)
- [Insights Glossary for Snapchat Creators](#)
- [How do I view and share replies to my Question sticker on Snapchat?](#)

Company	Community	Advertising	Legal
Snap Inc.	Snapchat Support	Snapchat Ads	Other Terms & Policies
Careers	Spectacles Support	Advertising Policies	Law Enforcement
News	Community Guidelines	Political Ads Library	Cookie Policy
Privacy and Safety		Brand Guidelines	Cookie Settings
		Promotions Rules	Report Infringement

Snap Inc.

- Privacy Policy
- Terms of Service

